DIETARY PATTERN TRAJECTORIES OVER TIME AND DIABETES AMONG CHINESE ADULTS

Carolina Batis Ruvalcaba

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Nutrition (Epidemiology), Gillings School of Global Public Health.

Chapel Hill
2013

Approved by:

Barry Popkin, PhD

Linda Adair, PhD

Penny Gordon-Larsen, PhD

Michelle Mendez, PhD

Daniela Sotres-Alvarez, DrPH

Abstract

CAROLINA BATIS RUVALCABA: Dietary Pattern Trajectories over Time and Diabetes among Chinese Adults.
(Under the direction of Barry Popkin)

Dietary patterns, instead of single nutrients or foods, are a useful approach to study diet and diet-disease associations. However, most studies examine dietary patterns only at one point in time. The purpose of this dissertation was to identify the longitudinal changes or stability of dietary patterns and their association with Diabetes in the China Health and Nutrition Survey from 1991 to 2009 (7 waves of diet data). Aim 1: we derived two dietary patterns using factor analysis in each wave: a traditional southern pattern (rice, vegetables, meat, poultry and fish) and a modern high-wheat pattern (wheat products, nuts, fruits, eggs, milk and instant noodles/frozen dumpling). The structure of these patterns remained stable over time, but the tracking was lower and the adherence increased over time for the modern high-wheat. Aim 2: among 4,316 adults not previously diagnosed with diabetes the adjusted Odds Ratio for diabetes prevalence in 2009, comparing the highest versus the lowest dietary pattern score quartile in 2006, was 1.25 (0.78, 2.01) for the modern high-wheat pattern, 0.79 (0.51, 1.21) for the traditional southern pattern and 2.36 (1.55, 3.58) for a pattern derived with Reduced Rank Regression (with HbA1c, HOMA-IR and glucose as response variables). This pattern combined items of the modern high-wheat pattern (wheat products and soy milk) with items opposite to the traditional southern (low rice, poultry and fish). Aim 3: A score for

the third dietary pattern was estimated for each subject at each wave and with Latent Class Trajectory Analysis subjects with similar trajectories of their dietary pattern's score over time were grouped in 5 classes. Among two classes with similar scores in 2006, the one with lower scores from 1991-2004, had significantly lower HbA1c [-1.64 (-3.17, -0.11)], and non-significantly lower odds of diabetes. All together our findings suggest that the popularity of a modern high-wheat pattern was increasing and that part of this pattern, when combined with low intake of rice, poultry, fish and legumes, was associated with diabetes. In addition, even if the diets were similar recently, the long-term trajectories of this dietary pattern were also associated.

Acknowledgments

I was really happy during my time here at UNC. I wish to thank not only those that participated directly in this dissertation, but also everyone involved in making these four years of my life so memorable.

To my advisor, Barry Popkin. He had faith on me and gave me the opportunity to be here. He identified my strengths and weaknesses as a researcher and guided me accordingly. He always put the right amount of pressure on me, enough to challenge me and make me improve, but not too much to make me feel overwhelmed or discouraged. He gave me freedom and allowed, supported and encouraged me to follow my research interests. I'm immensely grateful to him; he was always kind and caring.

To my committee members, Linda Adair, Penny Gordon-Larsen, Michelle Mendez and Daniela Sotres-Alvarez. With their different expertise and teaching styles they complemented each other. They had the right answers, and they had the right questions. I was always amazed by how much their comments, revisions, and feedback make me think through, re-think and clarify my own thinking.

To Shufa Du, Bonnie, Fei and Zhihong. They taught me about Chinese dietary culture and with their help I felt more comfortable studying this diet unfamiliar to me. I especially thank Shufa Du for getting me started with the China Health and Nutrition Survey and their data sets; and Bonnie for sitting with me during many hours going over the foods, food groups and my papers, making sure my treatment and interpretation of the data was accurate to the Chinese context.

To all the Nutrition and Epidemiology professors I interacted with. Through their carefully designed courses, they gave me the basis and helped me develop my critical thinking. Outside the classroom, through meetings or informal chats, they kept building my understanding of the concepts, shared their experiences, and offered advice for personal and professional development.

To all the Nutrition students. I was always surrounded by friends and felt welcomed at all times, I'm mostly grateful for the lively and collaborative environment created by all students. I especially thank my nutritional epidemiology cohort (aka "tres morenas") *Carmen and Niha*. They're smart and hardworking, but also centered, laidback, easygoing and caring. We went through everything together, and it was so much easier and fun because of them. I also want to thank my office neighbors *Jennifer, Dalia, Chris, Kevin and Beth*, the students I interacted with, at lunch and work, on a daily basis. They taught me we don't need decades of research experience to be able to figure things out and help each other; having someone like them interested in listening, asking and thinking aloud was all I needed most of the times.

To my "almost-all-Spanish-speaking"friends Laura, Kimon, Jorge, Daniela, Yani, Arianna, Mau, Dalia, Wendy, Wayne, Erik, Ceci and Tim. Full of fun activities and ready to help, they were my family at Chapel Hill. I shared with them all the important events such as Thanksgiving, Christmas, and the soccer games of the Mexican team.

To my parents. They were with me the whole time through Skype. They are really proud of me, but I know they would be proud of me with or without a PhD. All they care is that I'm happy and knowing that they love me just because I'm their daughter is what I appreciate the most.

To my best friend and husband, Germán. He was loving and supportive, and his company was key to my happiness during this experience. He kept my mind engaged, challenged and entertained with new points of view, topics, ideas and questions. He pushed me to learn about myself and my abilities. I wanted to pursue a PhD abroad, but I never thought that universities at the level of UNC were in my realistic range of possibilities. He was the one that said "why not?" and encouraged me to apply.

Table of Contents

List of Tables

List of Figures

List of Abbreviations and Symbols

BMI	Body Mass Index
CHNS	China Health and Nutrition Survey
FFQ	Food Frequency Questionnaires
HOMA-IR	Homeostasis Model of Insulin Resistance
INFS	Chinese Institute of Nutrition and Food Safety
LCTA	Latent class trajectory analysis
NA	Not applicable
PCA	Principal Component Analysis
RRR	Reduced Rank Regression
UNC-CH	University of North Carolina at Chapel Hill

Chapter 1. Introduction

Background

The burden of chronic diseases is increasing rapidly in China and around the world, and diet is an important determinant of these conditions. The study of data-driven dietary patterns can help enhance our understanding of human dietary practices and their relationship with chronic diseases such as obesity and other cardio-metabolic outcomes. An important limitation in this specific area of research is that most of the studies associating dietary patterns with chronic diseases are cross-sectional or prospective cohorts with dietary intake measured only at baseline. These studies assume that diet remains constant over time in all individuals, an assumption that might be incorrect particularly in populations with rapid changing environments.

Some studies have addressed this limitation by examining dietary patterns with repeated measures of diet over time; however none of these have been conducted in Chinese population. In addition, there are only few studies performed in different populations around the world that have assessed the effects of inter-individual changes in dietary patterns trajectories.

We proposed that the incorporation of techniques like Latent Class Trajectory Analysis (LCTA) to identify trajectories in dietary patterns over multiple measurement occasions is relevant to the field. By embracing and exploring the individual histories in diet we can further enhance our understanding of human dietary practices and it association with health outcomes.

We used The China Health and Nutrition Survey (CHNS), a large cohort that includes a highly diverse sample with variations in a wide-ranging set of socioeconomic factors (income, employment, education, and modernization) and related health, nutritional, and demographic measures. It has repeated detailed measurements of diet over the course of almost 20 years (1991, 1993, 1997, 2000, 2004, 2006, 2009). This data set offers unique opportunities to contribute to the dietary patterns research field; it has multiple repeated measures over a long period of time, in a rapid changing environment and in population seldom studied in this particular area.

Research Aims

Aim 1. Study the changes and stability of dietary patterns over 18 years in Chinese adult population.

1a. Derive the dietary patterns independently at each point in time to determine whether the structure of dietary patterns have changed or remained stable.

1b. Using applied scores assess the tracking, trends over time, and socio-demographic characteristics associated with the adherence to each dietary pattern.

When examining dietary patterns over time there are two different dimensions one can look at, one is the changes or stability in the general structure of the dietary patterns (are foods combined differently in each year?), and another is the individuals' level of adherence to each dietary pattern over time. To evaluate the individual's level of adherence over time it is necessary to use applied scores. The dietary patterns from one point in time are used to derive the pattern score at all other times. This way a score representing the same dietary pattern over time is obtained.

Aim 2. Identify the association between dietary patterns in 2006 and diabetes.

2a. Evaluate the association between the dietary patterns derived with Factor (FA) or Principal Component Analysis (PCA) and diabetes prevalence.

2b. Complement the analysis by comparing previously the derived identified patterns and their strength of association with a dietary pattern derived from Reduced Rank Regression (RRR) (HbA1c, fasting glucose and HOMA-IR as response variables).

The strength of one method is the limitation of the other; PCA identifies patterns that have public health relevance because they describe the actual dietary patterns of the population whereas the foods in the RRR are not necessarily consumed together and hence could be behaviorally irrelevant. Conversely, the patterns from RRR are by definition associated with the outcome or response variables, which might not be the case for PCA patterns. Therefore both methods can complement each other and provide useful insights when compared side by side.

Aim 3. Evaluate the association of different dietary pattern trajectories with diabetes.

3a. Identify groups of individuals with similar dietary pattern trajectories, and the socio-demographic characteristics associated with group membership.

3b. Estimate the association of the different dietary pattern trajectories with HbA1c, insulin resistance and diabetes prevalence.

Using the factor loadings of the dietary pattern in 2006 that had the strongest association with diabetes, we calculated an applied score for each individual from 1991 to 2006. Latent Class Trajectory Analysis was then used to group people according to the similarities in their trajectories. The membership to group trajectories was treated as the exposure to evaluate the association with diabetes.

Chapter 2. Literature Review

The burden of chronic diseases globally and in China

Globally the burden of chronic diseases is increasing rapidly. In 2001 it was estimated that 60% of all deaths were due to chronic diseases.[1] China is no exception to this world scenario, in fact it is one of the countries with the most rapid increase in obesity and other chronic diseases[2]. From 1989 to 1997 the prevalence of a BMI higher than $25kg/m^2$ tripled among men and doubled among women in China. [3] In addition to China's increased burden, it is particularly important to study the factors associated with chronic diseases in this population because 1) China is the most populated country in the world, with over 1.3 billion people, and any given prevalence of a chronic disease affects therefore an absolute significant number of persons. For example, 92 million of Chinese are diabetic[4] and there are more persons with cardiovascular disease in India and China than in all the economically developed countries in the world combined [1]. 2) Compared to Western populations, Asians have higher abdominal obesity and insulin resistance; and, they develop cardiometabolic risk factors at lower BMI and at younger ages [5-8]

The important role of diet and nutrition as determinants of chronic diseases is well established.[1] Therefore, continued research efforts should be done in order to better understand the determinants of dietary choices and its health consequences in the population.

Nutrition transition and dietary changes in China

Rapid economic and social changes have contributed to rapid shifts in the diets of the Chinese population. The traditional Chinese diet consisted of cereals and vegetables with few animal products, but a marked increase in the diversity of the diet has been observed.[9] The changes seen follow a Westernization pattern that is due to the introduction of a free market for food products. The intake of edible oils and animal-source foods has increased; e.g. animal products more than tripled from 1952 to 1992. On the other hand, cereal intake has decreased considerably during the last 2 decades, especially coarse cereals. Vegetable intake has remained mainly unchanged in urban areas but slightly decreased in rural areas; while fruits have slightly increased.[9-11] All of these changes have led to an increase in the overall energy density of the diets. The food marketing landscape has also changed considerably since China has the world's fastest growth in supermarkets. Fast-food chains are rapidly expanding in the cities and the intake of soft-drinks is starting to increase. However, Chinese people still consume limited amounts of Western-style fast foods and sweetened beverages compared to Western societies.[9-11]

Although the general changes over time of food groups consumption in the Chinese population has been clearly identified; the understanding of how these foods are eaten in combination, and the changes (or stability) in these combinations at the population and at the individual level is limited. This is especially important because while society changes some particular individuals might not do so. In order to improve policies and interventions, it is necessary to identify if several specific age, period or geographical cohorts remain with a constant diet while others change.

How does the study of dietary patterns contribute to the field of nutrition epidemiology?

Dietary patterns have been proven as a useful tool for studying diet. Contrary to the single nutrient or food group approach, dietary patterns consider multiple foods. And because people select foods in combination and not single nutrients, dietary patterns most closely resemble the actual eating behavior[12-14]. In addition, studying dietary patterns is advantageous to exploring associations of diet to disease. For instance, 1) the single nutrient approach may be inadequate for accounting for complicated interactions among nutrients, 2) the high correlation between nutrients could make difficult the investigation of their separate effects, 3) the effect of a single nutrient may be too small to detect, 4) analysis for multiple nutrients or food groups can yield statistical significant results simply by chance and 5) nutrient or foods are commonly associated with specific dietary patterns which may cause confounding hard to control for. [14]

In general dietary patterns can be grouped in 2 categories: theory-driven or data-driven. Theory-driven are score-based approaches based on dietary recommendations (i.e. Healthy Eating Index[15]). Data-driven or empirical dietary patterns use statistical methods to reduce many food or food groups into a set of meaningful patterns that describe the way people eat.

One benefit of data-driven patterns over dietary indexes is that they can help us understand better human dietary practice, because from data-driven patterns we learn which foods are eaten in combination. This can provide important insights for dietary interventions. In addition, by studying data-driven dietary patterns we are not evaluating the effects “ideal patterns" but the effect of patterns that are already followed by the population. Therefore, in this proposal we will only focus on data-driven dietary patterns.

There are two common approaches to derive data-driven dietary patterns: 1) factor analysis, which identifies linear combinations of foods that are frequently consumed together, and gives a summary score to each individual for each pattern; and 2) cluster analysis, where individuals are classified into mutually exclusive groups based on the similarities between their diets)[14,16]. In spite of the subjectivity involved during the process of dietary pattern derivation, consistent results have been found for the association with some diseases. The "western", "junk food", "cosmopolitan", "empty calories" patterns have been associated with increased risk, and "healthy" or "prudent" patterns high in vegetables, fruits, fish, reduced-fat dairy, cereals, have been associated with decreased risk.[16,17]

However, findings for BMI, obesity or waist circumference have been inconsistent. This was apparent in a literature review of studies examining the association between dietary patterns and BMI.[18] From nine studies that used factor analysis, three did not find a significant association. Among the ones that found significant associations, the types of patterns associated with lower BMI ranged from: "low culinary complexity" (confectionery, butter, cookies), "satiating" (macaroni, sausage, white bread), "convenience" (beer, chips, sauces), "high fat/sugar-dairy", "traditional" and "prudent". Whereas some of the patterns positively associated with BMI were "fruit", "high fat" (eggs, bacon, sausage), "salad", "bush food" (rabbit, duck, fish), "fruit juice" (fruit juice, high sugar drinks, canned fruit) and "western" (red meat, processed meat, refined grains). Most of the studies were cross-sectional, a reason that might largely explain the inconsistent results. Studies looking at longitudinal changes in anthropometry have found that patterns rich in reduced-fat dairy products and high-fiber foods are associated with smaller gains in weight and waist circumference[19]. And that increases in healthy pattern score between two points in time is associated with either decreases in BMI or smaller increases in BMI[20].

In terms of diabetes and diabetes related markers, several studies in populations around the world using factor analysis have been conducted. Healthy dietary patterns have been inversely associated with insulin resistance or insulin, whereas Western patterns (higher intakes of red meats, high-fat dairy products, and refined grains) have shown the opposite effect [21,22]. Similarly, patterns primarily rich in vegetables and fruit have been associated with reduced risk of incident diabetes. And patterns characterized by dim sum and meat; or red and processed meats, sweets and desserts, french fries, and refined grains; or butter, potatoes, and whole milk have been found to be associated with increased risk of incident type 2 diabetes.[23-26]

Despite the contributions of dietary patterns to the study of diet and diet-disease associations, an important limitation is that most of the studies measure diet only at one point. They assume that diet remains constant over time in all individuals, and that therefore one point in time represents long-term intake. This assumption might be incorrect particularly in populations with rapidly changing environments. In addition these studies also assume that one point captures the relevant time-frame for disease development, leaving out the possibility of testing different time-frames. This gap in the literature may even be a reason for the inconsistent results found for obesity outcomes.

Therefore the interest in dietary patterns over time has increased in the last decade. However comparisons over time are not straightforward because dietary patterns are population and time specific. Thus, so far only few studies have examined dietary patterns with repeated measures of diet.[20,27-33]

Analyses of dietary patterns over time

The studies of dietary patterns over time, conducted so far have had the following purposes:

1) Looked at the stability of dietary patterns in the population. One approach was to derive dietary patterns independently at each point in time and to assess if the patterns remain similar or if new patterns emerge. Another approach was to apply the scores or characteristics of the patterns found in one point to the rest of the measurements, so that the same patterns were analyzed over time[27,29,32,33]. Stability has been usually examined using correlation of scores or agreement between quantiles in 2 points in time.

2) Evaluated the within-subject difference in dietary patterns, by looking at the change in quantiles or score differences in 2 points in time, and analyzed how these differences affected health outcomes[20,30].

3) Examined the association between a dietary pattern and a heath outcome using multilevel models. These models used the repeated dietary patterns for obtaining an average dietary pattern-health outcome association while adjusting by the inter-individual correlation of repeated measures[28,31].

All of these studies have importantly contributed to our understanding of dietary practice over time. For example Cutler GJ et al. identified a new “fast-food” dietary pattern in adolescents at age 20 that was not observed when they were 16 years old. Whereas, in the teenagers that were examined at 13 and 17 years old, the patterns remained fairly constant at both times[32]. If this analysis was performed by looking only at the food groups level (instead of dietary patterns), the authors could had identified an increase in fast-foods at a certain age, but with this analysis they were able to identified that these foods were in addition eaten in combination. This study essentially told us that people select not just one food but major combinations and it is often misleading to examine one food when multiple foods are correlated.

No analyses on dietary patterns over time have been conducted in Chinese population. As exemplified above, we can gain meaningful insight and knowledge about the dietary practices in China by using this type of analysis. Moreover, not only no analyses over time have been conducted in this country, but also studies with only diet at one point in time are scarce. It is particularly important in the research field of dietary patterns to have population specific studies. It has been seen that even between different countries, both dietary patterns and their relationship with disease have similarities, however, important cultural differences might exist [14,17]. As an example, patterns rich in vegetables are generally associated with lower risk of obesity and chronic diseases; however a study among Chinese found that the vegetable-rich food pattern was associated with oil and energy intake and with increased risk of obesity [34]. This emphasizes the need of extending the research in dietary pattern in this region, which we propose to address in our study.

Another important gap in the study of dietary patterns over time is that most studies are limited to two points in time and do not explore individual trajectories. As described before, the studies done so far have been limited to the assessment of the stability or changes in dietary patterns between only 2 points in time. Even when more time points were included in the analysis simultaneously, the aim was not to characterize the individual trajectories but to adjust by them. These methods do not adequately represent the dynamic history of the dietary exposure, and therefore have limited ability to detect when the exposure has the highest effect on adverse health outcomes. A step beyond, would be to trace the individuals dietary trajectories with all available dietary data over a long period of time, and group individuals with similar trajectories to assess what socio-demographic and health outcomes they also share. Individuals may have a similar dietary pattern at one point in time, but some of them could have recently adopted this pattern and some could have followed the same pattern for years. It is essential to understand if the associated health risks for these two types of trajectories are different.

Therefore, in this research, we characterized the individual trajectories in dietary patterns over almost 20 years (7 points of repeated measures of diet). The technique that we incorporated in the field of dietary patterns with this purpose is LCTA. Conventional latent growth curve models or longitudinal mixed models assume that a single trajectory can adequately approximate the entire population, so they just a give a single trajectory average. The heterogeneity in these models is captured only by variations in the slope and intercepts (random effects). In contrast LCTA, applied to the trajectories of dietary patterns, could allow the identification of different groups or classes of trajectories within the population.[35,36] This technique has been used previously to characterize trajectories of BMI[37] or other social characteristics such as alcoholism,[38] depression,[39] delinquency,[40] etc, but have not being used for dietary patterns before. The goal of this method is to cluster individuals so that within a group the trajectories are more similar than between the groups.[35]

Advantages of using the China Health and Nutrition Survey (CHNS)

This survey has many unique characteristics that overcome some of the limitations on previous studies and can help us get important insights in the dietary pattern-disease relationships. For instance, the long follow-up of the rapid changing environment in China, offers a unique opportunity to determine the time period where diet is most relevant for disease development. If the diet of all individuals remains constant over time, it cannot be distinguished if the relevant exposure is the most recent and it happened to be the same of a longer follow-up period, or if the diet of the entire follow-up is the relevant time-frame.

The great heterogeneity in diet over time seen in this cohort, is also complemented with a wide temporal and spatial variation in socioeconomic factors such as income, employment, education and modernization. This variation can enabled us to better understand how do these factors affect dietary patterns.

In addition, the dietary data in the CHNS was carefully collected and has a high level of detail. Three 24hr recall were combined with a weighting of all foods and ingredients consumed in the household. This way, the use of standard recipes that may not capture individual intake differences was avoided. This methodology for assessing diet intake is particularly appropriate for the study of dietary patterns over long follow-up periods. Most studies have used Food Frequency Questionnaires (FFQ)[20,27,29,32,33], which may not be able to detect the introduction of new foods if the same instrument is used over time. And even if the FFQ instrument is updated, this raises methodological issues for comparing the intake over time.

Chapter 3. Longitudinal Analysis of Dietary Patterns in Chinese Adults from 1991 to 2009

Overview

Our aims were to identify the changes or stability in the structure of dietary patterns and the tracking, trends and factors related to the adherence of these patterns in China from 1991 to 2009. We used seven waves of the China Health and Nutrition Survey and included 9,253 adults with ≥3 waves complete. Diet was measured over a 3-day period with 24-hr recalls and a household food inventory. Using factor analysis in each wave we found that the structure of the two dietary patterns identified, remained stable over the studied period. The traditional southern pattern was characterized by high intake of rice, fresh leafy vegetables, low-fat red meat, pork, organ meats, poultry and fish/seafood and low intakes of wheat flour, corn/coarse grains; and the modern high-wheat pattern was characterized by high intake of wheat buns/breads, cakes/cookies/pastries, deep-fried wheat, nuts/seeds, starchy roots/tubers products, fruits, eggs/eggs products, soy milk, animal-based milk and instant noodles/frozen dumplings. Temporal tracking (maintenance of a relative position over time) was higher for the traditional southern, whereas adherence to the modern high-wheat had an upward trend over time. Higher income, education and urbanicity level were positively associated with both dietary patterns, but the association became smaller in the later years. These results suggest that even in the context of rapid economic changes in China; the way people chose to combine their foods remained relatively stable. However, the increasing popularity of the modern high-wheat pattern, a pattern associated with several energy-dense foods is cause of concern.

Introduction

China is one of the countries with the most rapid increases in chronic diseases; overweight/obesity (BMI $\geq 25 kg/m^2$) prevalence among adults was 15% in 1992, 22% in 2002[41] and 33% in 2007-08[42]. Diabetes, hypertension and dyslipidaemia reached a prevalence of 10%, 27% and 65% respectively in 2007-08[42]. Due to the key role that diet plays in all of these chronic diseases, it is important to better understand the eating behavior of this population.

Over the last two decades, many important changes in the diet of the Chinese population have been identified. Eating behaviors like snacking emerged and continue to grow, and cooking methods have also changed from predominantly steaming and boiling to frying[43,44]. A marked increase in dietary diversity has also been observed, with more people consuming food from a higher number of food groups[10]. The intake of vegetables, fruits, cakes, and milk and other animal products like pork, poultry and eggs have increased; whereas the intake of cereals and tubers has decreased[45].

In addition to the study of number of meals, cooking methods or food group consumption, data-driven dietary pattern analyses, such as factor or cluster analysis, are also useful for studying diet. Dietary patterns more closely resemble actual eating behaviors because they consider multiple food groups instead of single food groups or nutrients, and they give insights into how people eat by identifying the foods that are eaten in combination[12-14]. Previous studies conducted in a Chinese National survey from 2002 have identified several dietary patterns like the “Yellow earth” or “traditional northern” pattern that is high in wheat, wheat products, maize, sorghum and tubers; the “green water” or “traditional southern” pattern high in rice, vegetables, seafood, pork and poultry; and the “western” or “new affluence” pattern high in beef, lamb, milk, cheese, yogurt, cakes, juices and nuts[46-48]. However to best of our knowledge, no longitudinal analysis of dietary patterns over time has been conducted in the Chinese population.

To fill this gap, we used measurements of dietary intake from seven occasions over the course of 18 years collected in the China Health and Nutrition Survey (CHNS). When examining dietary patterns over time there are two different dimensions one can look at, one is the changes or stability in the general structure of the dietary patterns (are foods combined differently in each year?), and another is the individuals’ level of adherence to each dietary pattern over time. Therefore, our aims were to first derive the dietary patterns independently at each point in time to determine whether the structure of dietary patterns have changed or remained stable. Secondly, we assessed the tracking, trends over time, and socio-demographic characteristics associated with the adherence to each dietary pattern.

Methods

Study design and participants

The CHNS is an ongoing study with detailed income, employment, education, demographic, health, and diet information. The survey was designed to examine across space and time how economic and social changes are associated with a range of health behaviors. A multistage, random cluster process was used to draw the sample in 9 provinces. Survey protocols, instruments, and the process for obtaining informed consent for this study were approved by the institutional review committees of the University of North Carolina at Chapel Hill (UNC-CH) and the Chinese Institute of Nutrition and Food Safety (INFS), China

Center for Disease Control and Prevention. Participants provided their written, informed consent. Additional details about the CHNS data are provided elsewhere[49].

Surveys were conducted in 1989, 1991, 1993, 1997, 2000, 2004, 2006 and 2009. We used data from 1991 to 2009 because in 1989 only adults aged 20-45 years were included. All waves of the CHNS obtained identical clinical, dietary and anthropometric data from each household member. We included all adults aged 18 to 65 years old at any wave with at least 3 waves with complete dietary data (N=9,253); from these 20% had all 7 waves of diet complete, 50% had 5 or more, and 75% had 4 or more waves of diet complete.

Dietary assessment and food grouping

The dietary assessment in the CHNS is a combination of three consecutive 24-hour recalls at the individual level and a food inventory at household level performed over the same three day period. The three consecutive days were randomly allocated to start from Monday to Sunday. For the food inventory, all available foods at the household (purchased, stored or home produced) were measured on daily basis with Chinese balance (1991-1997) or digital scales (2000-2009). The changes in the household food inventory as well as the wastage were used to estimate total household food consumption. For the 24-hour recall, trained interviewers recorded the types, amounts, type of meal and place of consumption of all food items consumed. For dishes prepared at home the amount of each dish was estimated from the household food inventory, based on the proportion of each dished the person reported to have consumed.

The food groups included in our analysis were based on a food grouping system developed specifically for the CHNS by researchers from UNC-CH and INFS[10], this system separates foods into nutritional and behavioral meaningful food groups. We did not include alcoholic beverages, because it was mostly consumed by males, with very low consumption for females. Further description of the food group classification that we used can be found in Supplemental Table 3.1.

Statistical analysis

Because in each wave we included subjects that were 18 to 65 y old, it was possible for subjects to be excluded at certain waves (e.g. a subject 64 y old in 2000 would be excluded in 2004 and after, or a subject aged 14 y old in 1991 would be included only when he became 18 in 1993). This exclusion criterion resulted in younger subjects at the later waves. In addition, the proportion of subjects in the sample residing in the North region increased since 2000. This was because in 1997 a province in the North was unable to participate and a substitute province was included in the study; and in 2000 both the original and the substitute provinces were included again. Therefore, all analyses were adjusted by geographical region and age in 1991(equivalent to adjusting by birth year), so that the dietary trends found over time were not related to these sample distribution changes in age and region.

Most food groups had a high proportion of non-consumers; possibly due to the fact that diet intake was measured over a 3-day period. Therefore, we categorized consumption as binary (non-consumers vs. consumers) for food groups with <25% of consumers in all waves; and otherwise as a three-level variable (non-consumers, consumers below or above the estimated median using all waves).

We performed exploratory factor analysis for categorical variables using the robust weighted least square estimator in Mplus 6.1 (Muthén & Muthén, Los Angeles, California) at each wave. Factor analysis on ordinal variables is performed with a polychoric correlation matrix. Conceptually, ordinal variables have an underlying continuous normally distributed variable, so thresholds for the levels of the categorical variables are estimated. These thresholds are normal z scores corresponding to the cumulative proportion of subjects in each category [i.e. for a binary variable (non-consumers vs. consumers) with a 0.87 proportion of non-consumers, the threshold between the two categories would be equal to the corresponding normal z-score 1.13].[50,51]

Food groups with ≤5% of consumers in all waves were not included in the factor analysis. These food groups are uninformative because of their intake's lack of variability in the population, and also they could produce bivariate tables with empty cells and affect the polychoric correlation.

Based on eigenvalues >1, inspection of the scree plot, and interpretability, we retained two dietary patterns in each wave. Factors were rotated with the varimax procedure, because this method seeks to maximize the variability among the loadings in each factor and hence gives simpler and more easily interpreted factors[52]. Because we found similarity in dietary pattern structure and factor loadings across all waves, we computed applied scores as many studies with longitudinal dietary patterns have done previously.[27,28,53,54] The loadings of all food groups in year 2000 were used to calculate the factor scores in all other waves; because this applied score measures the same dietary pattern across time, it allows meaningful temporal comparisons. We chose year 2000 because it was the midpoint of the observed period, the loadings were close to the mean of the loadings of all other wave, and therefore most representative of the entire period. In order to estimate the scores with the procedure implemented by Mplus 6.1 (similar to the regression method, but with an iterative technique for categorical variables[55]) we performed a confirmatory factor analysis in each wave specifying the factor loadings and thresholds estimated from year 2000. Because thresholds are z scores of the intake variables and they are involved in the estimation of the score, it is important that they are not year-specific; therefore we also fixed them to the estimates of year 2000.

Pearson correlation coefficients of the factor scores between each pair of waves were computed to assesses the tracking (maintenance of a relative position or rank over time) of each dietary pattern. Factor score means over time were computed, to look at trends in each dietary pattern. Finally, multiple linear regressions with each factor score as the outcome and socio-demographic variables as the predictors were performed independently in each wave, the clustering at the household level was accounted for in the estimation of the variance. Except for the factor analysis, all other analyses were conducted in Stata 12.1 (StataCorp, College Station, TX).

Sensitivity Analysis

In order to corroborate that the dietary patterns obtained at each wave were not affected by the sample changes in geographical region and age, we computed inverse probability weights and included them in the factor analysis. We fitted two logistic regressions, one to predict the probability of being in each wave, and another to predict the probability of being in each wave conditional on region and age in 1991. Then, stabilized weights were estimated as the ratio of these two probabilities. Applying these weights is an alternative to standardization[56]. We found that the dietary patterns were very similar to the ones we found in our original analysis; the difference between the factor loadings of the two analyses was below 0.08 in all food groups.

In addition, because subjects coming from the same households were more likely to consume the same type of food groups, we repeated the analysis using only a single member per household (selected randomly, n=4,837). We found that the dietary patterns did not change meaningfully; all loadings had a difference below 0.08. Also, the correlation coefficients between scores remained basically unchanged (all differences in the coefficients were below 0.03).

Results

There was a dramatic increase in the prevalence of overweight from 1991 to 2009 and in the proportion of the sample classified as medium and high income over the same period, which paralleled changes in urbanization. The proportion of smokers declined over time in males, and remained low in females (Table 3.1).

The mean total energy intake increased by 267 kJ/d (64 kcal/d) from 1991 to 2009 (Table 3.2). In general, it can be seen that the diversity of diet increased over time, the mean number of food groups consumed increased over time and for most of the food groups the percentage of consumers also increased. The only food groups with a decline in the percentage of consumers from 1991 to 2009 were: wheat flour, dried legumes, pickled/salted/canned vegetables, and low-fat red meat. In contrast, starchy roots/tubers products, fresh leafy vegetables, dried vegetables and organ meats remained with a stable percentage of consumers over time (difference ≤ 1 point between 1991 and 2009). In addition, all food groups that had an increase in the percentage of consumers, also had an increase in the amount consumed (g/per capita), except for rice, corn/coarse grains and starchy roots/tubers where even if the percent consumers increased, the g/capita actually decreased over time (data not shown).

The following food groups had ≤5% of consumers in all waves (data not shown): deep-fried rice/legumes, dried fruit, preserved fruit with added sugar, seaweed, processed meats, dairy products, sweetened dairy products, Western-style fast-food, salty snacks, ready-to-eat cereals/porridge, calorically-sweetened beverages, and low-caloric beverages. So even if these modern-type foods are gaining popularity in China, over a 3-day period, they were not yet widely consumed by 2009.

Using factor analysis on a total of 29 food groups (excluding food groups with ≤5% of consumers) we identified two dietary patterns, which we called "traditional southern" and "modern high-wheat". These dietary patterns, despite the increase in the diversity of diet, remained relatively stable (Table 3.3). Across time, the traditional southern dietary pattern was positively associated with the intake of rice, fresh leafy vegetables, low-fat red meat, low- and high-fat pork, organ meats, poultry and fish/seafood and inversely associated with wheat flour, corn/coarse grains. Interestingly, the loadings of rice, wheat flour and corn/coarse grains markedly declined over time, which was not related to the percentage of consumers (as seen above, of these food groups only wheat flour had a decrease in the percentage of consumers). However it can interpreted from these shifts in the loadings that there was a decreased in the influence or importance that these food groups had in this dietary pattern over time. The second dietary pattern, modern high-wheat was positively associated, across time, with the intake of wheat buns/breads, cakes/cookies/pastries, deep-fried wheat, nuts/seeds, starchy roots/tubers products, fruits, eggs/eggs products, soy milk and animal-based milk. In the earlier years this pattern was associated with high-fat red meat, high-fat pork, organ meats and poultry/game. In 1991 and 1993 there were no consumers of instant noodles/frozen dumplings, so this food group was not included in the factor analysis in these years, but when these foods began to be consumed in 1997, they became prominent in the modern high-wheat dietary pattern.

Because overall we found similarity in dietary pattern structure over time, we computed applied scores and hence the rest of the results refer to these applied scores. The correlation coefficients for scores between years were considerably higher for the traditional southern pattern compared to the modern high-wheat pattern; 0.67-0.80 vs. 0.46-0.63 respectively (Table 3.4). This means that the tracking of the traditional southern pattern was higher, or in other words that individuals maintained their relative position over time. The tracking of the modern high-wheat pattern was lower, representing that subjects were less stable in their relative position.

The mean dietary pattern score increased over time for both dietary patterns, however the slope was considerably higher for the modern high-wheat pattern (Figure 3.1 A). To understand how much of the increase in both scores was related to the increase in diet diversity, we adjusted the scores by number of food groups consumed (Figure 3. 1B). After adjustment, the traditional southern pattern had a flat trend over time, and the modern high-wheat had only a slight increase. This means that for the modern high-wheat pattern, in order to increase the score while holding the number of food groups constant, individuals were either substituting other foods with food groups related to this pattern or had a higher consumption of these.

To examine how the trends in the scores were translated into changes in dietary intake, we estimated the percentage of consumers among those in the 4 highest quartile of the scores of each dietary pattern in each wave (Table 3.5). The trends in consumption seen in the entire sample (i.e. increases in most of the food groups) are equally found in subjects with high scores for either pattern. However, the increase in percentage of consumers in the food groups positively related the modern high-wheat pattern, among the followers of this pattern, is dramatic. For example from 1991 to 2009 wheat buns/breads increased from 7 to 67%; cakes/cookies/pastries, 8 to 24%; deep-fried wheat 23 to 51%; fruits, 25 to 55%; soy milk, 6 to 44%; and instant noodles/frozen dumplings, 0 to 36%. This large increase in consumption of key food groups associated with the modern high-wheat dietary pattern can explain the upward trend observed in the score of this pattern. While we only present food groups relevant for the dietary patterns in Table 3.5, we also examined food groups excluded from the dietary pattern analysis due to low proportion of consumers (not shown), we found that none of these food groups had a higher percent of consumers in one of the dietary patterns, with the exception of deep-fried rice/legumes (traditional southern: 5-6% of consumers in the 4^{th} highest quartile, versus 1-2% of consumers in the modern high-wheat dietary pattern).

Finally, we run multiple linear regressions in each wave and looked at how socio-demographic factors were related to the traditional southern and modern high-wheat dietary patterns (Table 3.6). Geographical region was strongly associated with the traditional southern pattern, with lower scores in the central and north compared to the south. Interestingly the score difference between regions decreased over time, mainly due a slight increase over time in the North and Central region and a slight decrease in the score in the South. Urbanicity was strongly associated with the modern high-wheat pattern, with higher scores among those living in more urbanized areas. Males had a slightly higher score for traditional southern pattern and slightly lower score for modern high-wheat compared to females. Higher education, income, urbancity level, and alcohol intake were associated with higher scores in both dietary patterns, however the strength of the association decreased over time. Adjusting by number of food groups consumed did not affect the estimates, although differences for education, income and urbanicity level were largely attenuated (data not shown).

Discussion

In this longitudinal study that included 7 waves of diet data over the course of 18 years (from 1991 to 2009), we derived two dietary patterns using factor analysis: a traditional southern pattern characterized by rice, vegetables, meat, poultry and fish; and a modern high-wheat pattern characterized by wheat products, nuts, fruits, eggs, milk and instant noodles/frozen dumpling. Despite the rapid increase in diet diversity, the structure of these two dietary patterns remained stable over time, meaning that the type of foods the population chose to eat in combination has not changed. However, compared to the traditional southern pattern, the scores of the modern high-wheat pattern had lower tracking and steeper upward trend over time, indicating that this pattern is more dynamic and is becoming more popular over time. Higher education, income and urbanicity level were all related to higher scores in both patterns, but the difference in pattern scores between these socio-economic groups became smaller over time, reflecting that access to these dietary pattern is becoming more widely available.

Although this is the first study to assess longitudinal trends in dietary patterns over time in China, previous studies have looked at dietary patterns at one point. Whereas our findings differ from studies in specific urban areas, such as studies in Shanghai[57-60], they are most similar to findings in nationally representative samples, such as the China National Nutrition and Health Survey. Previously identified patterns are “yellow earth” or “traditional northern” high in wheat products, maize or sorghum and tubers; “green water” or “traditional southern" high in rice, vegetables, seafood, pork and poultry; and "Western" high in beef, fruit, eggs, poultry, seafood, tofu, milk, cake, fruit juice, beverages and nuts.[47,48] Our "traditional southern" pattern is comparable to the one previously described. However, our "modern high-wheat" pattern could be considered as a combination of the "traditional northern" and "Western" patterns that these previous studies have found. This is reasonable because even in these studies, a considerably high intake of wheat, fried-wheat and other cereals was observed in the "Western" pattern. In addition, in the Jiangsu province a pattern similar to our "modern high-wheat" was found, in which the western-type pattern was also related to fried-wheat and cakes[61]. Taken all together, it is clear that the regional tradition and availability of wheat in the North and rice in the South is still a main driver in Chinese's food selection. It is interesting that wheat-patterns are more likely characterized by intake of more varied and western-type foods. Possibly, compared to rice, there is a wider range of foods that can be prepared with wheat, and a wider range of foods that can be accompanying a wheat-based food, therefore facilitating the incorporation of new foods among subjects used to consume wheat.

Even if the general structure of both dietary patterns remained stable over time, the tracking or in other words the stability of the relative position of each subject's factor score between two points in time was higher for the traditional southern pattern than for the modern high-wheat. This is comparable to a study in Japan[62], where among the prudent, traditional and westernized patterns, the traditional had the highest tracking among men. The degree of tracking that we found in the traditional southern pattern was remarkably high, the correlation coefficient between scores from 18 years apart (1991-2009) was 0.68, which is similar to what other studies in the US[63], Sweden[64] and Japan[62] have found but in only one year apart (0.56 to 0.77). In studies with 4 or 10 years apart in England[53] and Sweden[29], correlation coefficients have been found to be around 0.30 to 0.52. The correlation coefficients in these studies, might not be directly comparable to ours because they, instead of applied, used natural scores (scores from each point in time are calculated from the loadings of the pattern specific to each point in time); however, previous studies using both applied and natural scores, have found very similar correlation coefficients[27,53]. This means that relative to dietary patterns in other countries, the traditional southern pattern is China is very stable and well defined, and subjects that have high scores at present are expected to maintain their high scores for a long period. However, it is noteworthy that the tracking in the traditional southern pattern decreased over time, between 1991 and 1993 the correlation coefficient was 0.79, whereas in more recent years, between a similar 2 or 3 years interval the correlations were lower (0.71 for 2004-2006 and 0.67 for 2006-2009), reflecting that with time, subjects might be losing their stable intake of this pattern.

In this study we confirmed that the increase in diet diversity previously observed in 1989 to 1997[10], is a phenomenon still going on through 2009. Factor scores would be affected by diet diversity because they are similar to a weighted sum of several food groups, so with the exception of food groups with negative loadings, it is expected that the higher the number of food groups consumed, to a lesser or greater extent the higher the factor score. Therefore in the context of increasing diet diversity, the absolute factor scores trend over time represents the degree of adherence to a certain pattern as well as the overall number of food groups consumed. An alternative is to adjust by number of food groups consumed, because the only way to increase the factor score while holding the number of food groups consumed constant is to substitute with food groups related to the diet pattern and/or to increase the amount consumed of these food groups. In our results we found that in comparison with the traditional southern pattern, the modern high-wheat had a higher increase over time, and that even after adjusting by number of food groups consumed, the modern high-wheat pattern showed an upward trend. A clear illustration of this trend is the remarkable changes in the diets of those following this pattern (subjects with scores in the highest quartile of their year); although the increase in food groups such as fruits and soy-milk is beneficial, the sharp increase in items like cakes/cookies/pastries, deep-fried wheat or instant noodles is worrisome. In addition, it is of public health importance that with time, this pattern seems to be becoming more available to all urban/rural, income, and education groups.

The key strength of our study is the several repeated measures of diet in the same subjects, which allows us to assess for the first time the stability and tracking of dietary patterns in the Chinese population. Even though there were changes in our sample because not all the subjects had the 7 waves of diet complete, we believe our comparisons between years were not affected by this, not only because all subjects were present for at least three waves but also because we adjusted all of our results by age and geographical region.

A limitation in our study was that the diet measurement only captured three consecutive days of intake. Though the advantages of this method are important because the diet measurement is very detailed and precise, and the same method can remained constant over time, a disadvantage for this analysis is that the intake of important food groups that in this country are still only episodically consumed could not be captured. Items like dairy products, candies/other high-sugar foods, western-style fast food, salty snacks or calorically-sweetened beverages had a very low proportion of consumers in a 3-day period and hence we were not able to include these in our factor analysis. Therefore, we do not exclude the possibility that a new, separate western pattern emerged within the studied period, and we were not able to identify it.

In line with this, another important limitation common to all countries, is that the number of foods available in the food supply exceeds by far the number of those available in food composition tables. In the US for example, there are over 85,000 uniquely formulated products in the food system, whereas national food composition tables only have around 7,600 unique foods[65]. China has the world's fastest growth in supermarkets[66], and our survey, because of the inability to keep up with the rapidly changing food supply landscape, is unable to capture all the changes in intake, particularly of processed packaged foods. In addition, our dietary assessment, focuses on the measurement of foods at the ingredient and not at the dish level, this is very advantageous because recipes vary considerably between households[10] and few mixed dishes are available in the Chinese food composition table, however we miss this important behavioral aspect of diet, and important changes in dish selection might be going on.

In sum, to the best of our knowledge, this is the first study to assess the stability and tracking over time of dietary patterns in the Chinese population. We found that the way foods are eaten in combination has not changed much between 1991 and 2009. The degree of adherence to a traditional southern pattern has remained unchanged and there was remarkable inter-individual consistency in consumption of foods central to this pattern. On the other hand, the modern high-wheat pattern has become more prominent in the population and there was more within-subject variation in their adherence level. The trend observed in this modern high-wheat pattern may as well reflect the global influence and economic rise of the last decades in China. Close attention should be given to this pattern because it is associated with many energy-dense foods that may affect diet quality. So far, the northern part of the country and subjects with better socio-economic position or in urban areas are more prone to follow this pattern, but it seems that the reach of this pattern is extending towards the general population. Among the followers of this modern high-wheat pattern, a willingness to diversify dietary intake and increase the intake of equally healthful foods like fruits was evident; therefore great opportunity also lies in the promotion of healthy foods and in the efforts to increase their intake in this population.

Tables and Figures

Table 3.1 General characteristics of study sample by wave*

	1991	1993	1997	2000	2004	2006	2009
N	5521	5732	6702	7410	6831	6488	5658
Mean Age in 1991, years	37.3	36.8	35.0	33.5	32.2	31.7	30.7
Geographical region†, %							
North	9.6	9.3	9.9	19.7	21.5	22.1	21.9
Central	37.2	37.4	36.9	33.4	32.8	32.9	33.1
South	53.2	53.3	53.1	46.9	45.7	45.0	45.0
Male, %	50.2	49.8	49.7	49.1	48.5	47.8	48.3
Overweight (BMI ≥23 kg/m^2 [67]), %	32.9	35.4	43.2	50.6	52.4	54.3	56.7
Education, %							
None	25.5	24.0	23.7	20.0	18.0	23.1	24.4
Primary school	23.7	23.7	24.7	24.4	25.8	20.1	21.9
≥Lower middle school	50.8	52.1	51.3	55.2	56.0	56.6	53.6
Income‡, %							
Low	87.8	82.2	74.5	63.6	56.0	50.5	33.6
Medium	7.5	13.0	19.8	27.9	28.9	30.2	33.2
High	0.6	1.2	2.5	5.9	14.2	18.6	33.4
Urbanicity §, %							
Low	72.8	67.9	57.6	49.3	45.9	40.7	32.7
Medium	25.2	30.6	32.1	26.6	22.7	27.0	34.0
High	2.1	1.5	10.4	24.1	31.5	32.3	33.3
Currently smoking, %							
Females	3.1	3.5	3.8	3.3	3.2	2.9	3.3
Males	69.8	67.9	64.5	63.6	60.8	58.5	58.6
Alcohol intake ≥3 times/week, %							
Females	1.7	2.1	2.9	2.7	1.9	1.9	2.4
Males	25.6	29.9	30.2	32.1	31.9	31.7	31.7

*All variables (except age in 1991 and geographical region) are adjusted by age in 1991 and geographical region.

†Estimated from per capita household income inflated to 2009, categories based on cutoff values of tertiles in 2009.

‡North: Heilongjiang and Liaoning provinces; Central: Shandong, Henan and Jiangsu; South: Hubei, Guizhou, Hunan and Guangxi.

§ Estimated from urbanicity index, a multicomponent scale that considers population density, economic activity, modern markets, transportation, etc.[68], categories based on cutoff values of tertiles in 2009.

Table 3.2 Mean energy intake, number of food groups consumed and percentage of consumers for different food groups by wave*

	1991	1993	1997	2000	2004	2006	2009
Energy, kJ/day†	6586	6548	6431	6607	6623	6745	6853
Number of food groups consumed, (min, max)	7.2 (1, 19)	7.5 (2, 19)	7.9 (2, 21)	8.2 (2, 19)	8.7 (2, 20)	9.1 (2, 22)	9.8 (2, 21)
Percentage of consumers, %							
Rice	84	84	87	88	87	87	90
Wheat noodles	23	32	38	42	47	54	55
Wheat flour	51	42	40	41	36	28	25
Wheat buns, breads	3	12	13	15	28	34	37
Cakes, cookies and pastries	3	4	3	3	6	8	8
Deep-fried wheat	6	8	7	8	12	14	15
Corn and coarse grains	21	23	22	22	24	23	26
Starchy roots and tubers	41	39	41	37	44	45	50
Fresh legumes	38	37	41	47	47	54	53
Dried legumes	14	14	13	12	13	12	12
Legume products	41	41	50	49	50	49	54
Nuts and seeds	8	6	8	9	9	8	11
Starchy roots and tubers products	9	9	11	11	11	8	9
Fresh vegetables, non-leafy	75	77	76	83	84	87	90
Fresh vegetables, leafy	86	87	87	86	88	83	85
Pickled, salted or canned vegetables	30	32	26	25	22	22	23
Dried vegetables	9	11	9	9	7	8	10
Fruits	11	10	12	11	17	23	32
Low-fat red meat	6	9	7	7	4	3	4
High-fat red meat	4	5	8	8	12	13	10
Low-fat pork	9	11	11	10	12	12	13
High-fat pork	55	57	57	64	64	69	73
Organ meats	9	9	9	10	9	8	10
Poultry and game	11	12	16	18	18	18	24
Eggs and eggs products	34	32	49	51	53	59	64
Fish and seafood	32	32	34	34	36	37	42
Soy milk	2	2	2	4	5	10	12
Animal-based milk	1	1	1	3	7	6	5
Instant noodles and frozen dumplings	0	0	2	2	4	12	15

*All variables are adjusted by age in 1991 and region, food groups with ≤5% of consumers in all waves not shown.

† conversion factor 1 kcal = 4·184 kJ

Table 3.3 Exploratory Factor Analysis at each wave*

	Traditional southern dietary pattern							Modern high-wheat dietary pattern						
	1991	1993	1997	2000	2004	2006	2009	1991	1993	1997	2000	2004	2006	2009
Rice	**0.87**	**0.87**	**0.73**	**0.56**	**0.76**	**0.64**	**0.50**	—	-0.25	-0.25	**-0.41**	-0.20	-0.28	**-0.30**
Wheat noodles	**0.30**	0.27	0.28	—	—	—	—	—	—	—	—	—	—	-0.22
Wheat flour	**-0.80**	**-0.82**	**-0.86**	**-0.66**	**-0.59**	**-0.67**	**-0.50**	—	—	—	0.26	—	—	—
Wheat buns, breads	—	—	**-0.30**	-0.27	**-0.53**	**-0.44**	**-0.49**	**0.33**	**0.40**	**0.39**	**0.40**	**0.37**	**0.48**	**0.45**
Cakes, cookies and pastries	—	—	—	—	—	—	—	**0.31**	**0.48**	**0.53**	**0.51**	**0.57**	**0.53**	**0.41**
Deep-fried wheat	—	-0.28	—	—	**-0.37**	—	—	**0.46**	**0.70**	**0.57**	**0.71**	**0.63**	**0.76**	**0.74**
Corn and coarse grain	**-0.84**	**-0.78**	**-0.76**	**-0.65**	**-0.64**	**-0.53**	**-0.35**	—	—	—	0.28	—	—	0.28
Starchy roots and tubers	—	—	—	—	—	—	-0.27	—	—	—	—	—	0.20	—
Fresh legumes	—	—	—	—	—	—	-0.20	—	—	—	—	—	0.20	—
Dried legumes	—	—	—	—	—	—	0.23	—	—	—	—	—	—	—
Legume products	—	—	—	—	—	—	—	0.24	—	—	—	—	—	—
Nuts and seeds	—	—	—	—	—	—	0.21	**0.42**	**0.33**	0.24	0.26	0.28	**0.34**	**0.30**
Starchy roots and tubers products	**-0.30**	—	—	—	—	—	—	**0.30**	**0.32**	0.26	0.26	0.24	0.28	0.26
Fresh vegetables, non-leafy	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Fresh vegetables, leafy	0.25	0.21	0.22	0.22	0.21	**0.32**	**0.46**	—	—	0.23	—	—	—	—
Pickled, salted or canned vegetables	0.28	—	0.21	0.25	—	—	—	—	—	—	—	—	—	—
Dried vegetables	—	—	0.21	—	**0.31**	0.24	**0.33**	—	—	0.25	—	0.25	—	—
Fruits	—	—	—	—	—	—	—	**0.43**	**0.57**	**0.46**	**0.35**	**0.56**	**0.55**	**0.44**
Low-fat red meat	**0.31**	**0.50**	**0.40**	**0.48**	**0.33**	**0.47**	**0.38**	**0.40**	0.25	0.28	0.20	—	—	—
High-fat red meat	—	—	—	0.24	—	**0.28**	**0.27**	**0.37**	**0.37**	**0.46**	0.24	**0.36**	0.22	0.29
Low-fat pork	**0.32**	**0.40**	0.26	**0.32**	**0.49**	**0.30**	**0.35**	0.25	**0.30**	—	—	—	—	—
High-fat pork	**0.32**	**0.36**	**0.44**	**0.48**	0.27	**0.34**	**0.43**	**0.49**	**0.34**	**0.44**	0.25	**0.32**	—	—
Organ meats	**0.40**	**0.49**	**0.39**	**0.53**	**0.32**	**0.43**	**0.48**	**0.45**	**0.32**	**0.46**	0.24	**0.37**	—	—
Poultry and game	**0.31**	**0.40**	**0.32**	**0.53**	**0.34**	**0.49**	**0.48**	**0.44**	**0.44**	**0.47**	0.25	**0.40**	—	—

Eggs and eggs products	–	–	–	–	–	–	–	**0.48**	**0.45**	**0.42**	**0.38**	**0.39**	**0.37**	**0.30**
Fish and seafood	**0.34**	**0.39**	**0.37**	**0.48**	**0.45**	**0.46**	**0.43**	**0.44**	**0.33**	**0.41**	–	**0.34**	0.26	0.24
Soy milk	–	–	–	–	-0.27	–	–	**0.55**	**0.79**	**0.58**	**0.72**	**0.69**	**0.76**	**0.78**
Animal-based milk	–	–	–	–	–	–	–	**0.64**	**0.68**	**0.56**	**0.58**	**0.67**	**0.53**	**0.49**
Instant noodles and frozen dumplings	NA	NA	–	–	–	–	–	NA	NA	**0.47**	**0.40**	0.27	**0.41**	**0.44**

NA; not applicable

[a]Loadings >=0.30 are in bold numbers, loadings <0.20 not shown.

Table 3.4 Tracking of dietary patterns: Pearson correlation coefficients of factor scores between waves for traditional southern (lower diagonal) and modern high-wheat (upper diagonal).*

	1991	1993	1997	2000	2004	2006	2009
1991		0.63	0.55	0.52	0.52	0.49	0.46
1993	0.80		0.61	0.56	0.57	0.52	0.48
1997	0.79	0.81		0.57	0.56	0.52	0.48
2000	0.73	0.76	0.76		0.56	0.53	0.48
2004	0.74	0.77	0.74	0.72		0.62	0.56
2006	0.73	0.76	0.72	0.70	0.72		0.59
2009	0.69	0.72	0.71	0.68	0.67	0.67	

*Pearson correlation coefficients were estimated between all available subjects for each pair of waves, therefore sample size for each coefficient is different. Sample size ranges from n=2888 (1991-2009) to n=5884 (2000-2004)

Table 3.5 Mean energy intake, number of food groups consumed and percentage of consumers for relevant food groups, among those in the 4th highest quartile of the factor score in each dietary pattern and year.*

	Traditional southern			Modern high-wheat		
	1991	2000	2009	1991	2000	2009
Energy, kJ/day†	6947	7272	7430	7158	7140	7667
Number of food groups consumed	8.4	9.4	10.7	9.1	10.9	12.9
Percentage of consumers, %						
Rice	**100**	**100**	**100**	67	82	84
Wheat flour	**9**	**7**	**4**	82	61	35
Wheat buns, breads	3	5	15	**7**	**33**	**67**
Cakes, cookies and pastries	3	3	9	**8**	**12**	**24**
Deep-fried wheat	5	4	9	**23**	**31**	**51**
Corn and coarse grains	**1**	**2**	**4**	38	37	42
Nuts and seeds	10	12	16	**17**	**17**	**22**
Starchy roots and tubers products	6	11	7	**21**	**19**	**16**
Fresh vegetables, leafy	**97**	**95**	**96**	87	88	88
Fruits	13	19	40	**25**	**26**	**55**
Low-fat red meat	**16**	**19**	**9**	9	12	6
High-fat red meat	5	15	18	**9**	**16**	**18**
Low-fat pork	**18**	**20**	**24**	8	11	15
High-fat pork	**87**	**91**	**93**	75	82	83
Organ meats	**27**	**29**	**25**	13	19	14
Poultry and game	**30**	**44**	**50**	18	31	33
Eggs and eggs products	37	55	64	**57**	**77**	**83**
Fish and seafood	**60**	**65**	**68**	42	48	56
Soy milk	1	2	8	**6**	**15**	**44**
Animal-based milk	2	3	7	**4**	**11**	**18**
Instant noodles and frozen dumplings	0	2	14	**0**	**8**	**36**

* All variables are adjusted by age in 1991 and region, numbers in bold are for relevant food groups in each dietary pattern (loadings ≥0.20 in all waves), food groups not relevant for any dietary pattern are not shown.

† conversion factor 1 kcal = 4·184 kJ

Table 3.6 Factor score differences by sample characteristics in each wave

	Traditional southern			Modern high-wheat		
	1991	2000	2009	1991	2000	2009
Age in 1991, 10 years increment	0.01	-0.02*	-0.04*	0.02*	0.01	-0.02
Region						
Central vs. South	-1.07*	-0.83*	-0.77*	0.61*	0.72*	0.67*
North vs. South	-0.85*	-0.78*	-0.65*	0.60*	0.38*	0.60*
Male vs. female	0.03	0.07*	0.05*	-0.07*	-0.06*	-0.08*
Education						
Primary school vs. none	0.13*	0.11*	0.06*	0.11*	0.08*	0.06*
≥Lower middle school vs. none	0.15*	0.09*	0.04	0.23*	0.20*	0.19*
Income						
Medium vs. Low	0.27*	0.24*	0.11*	0.06*	0.11*	0.05
High vs. Low	0.17	0.35*	0.22*	0.22*	0.18*	0.18*
Urbanicity						
Medium vs. Low	0.26*	0.24*	0.20*	0.49*	0.38*	0.38*
High vs. Low	0.08	0.45*	0.43*	0.78*	0.61*	0.72*
Smoking vs. non-smoking	-0.05	-0.03	0.02	0.02	-0.01	0.01
Alcohol intake						
≥3 times/wk vs. <3 times/wk	0.16*	0.08*	0.06*	0.08*	0.10*	0.06*
N	4,785	6,172	5,553	4,785	6,172	5,553
Score mean	-0.14	0.00	0.11	-0.18	0.06	0.49
Score range (min, max)	-2.3, 2.4	-2.4, 2.3	-2.5, 2.2	-1.5, 2.3	-1.6, 2.6	-1.6, 3.0
Standard deviation	0.90	0.80	0.68	0.69	0.73	0.81

* $p<0.05$; score differences based on a multiple linear regression that included all variables shown in table

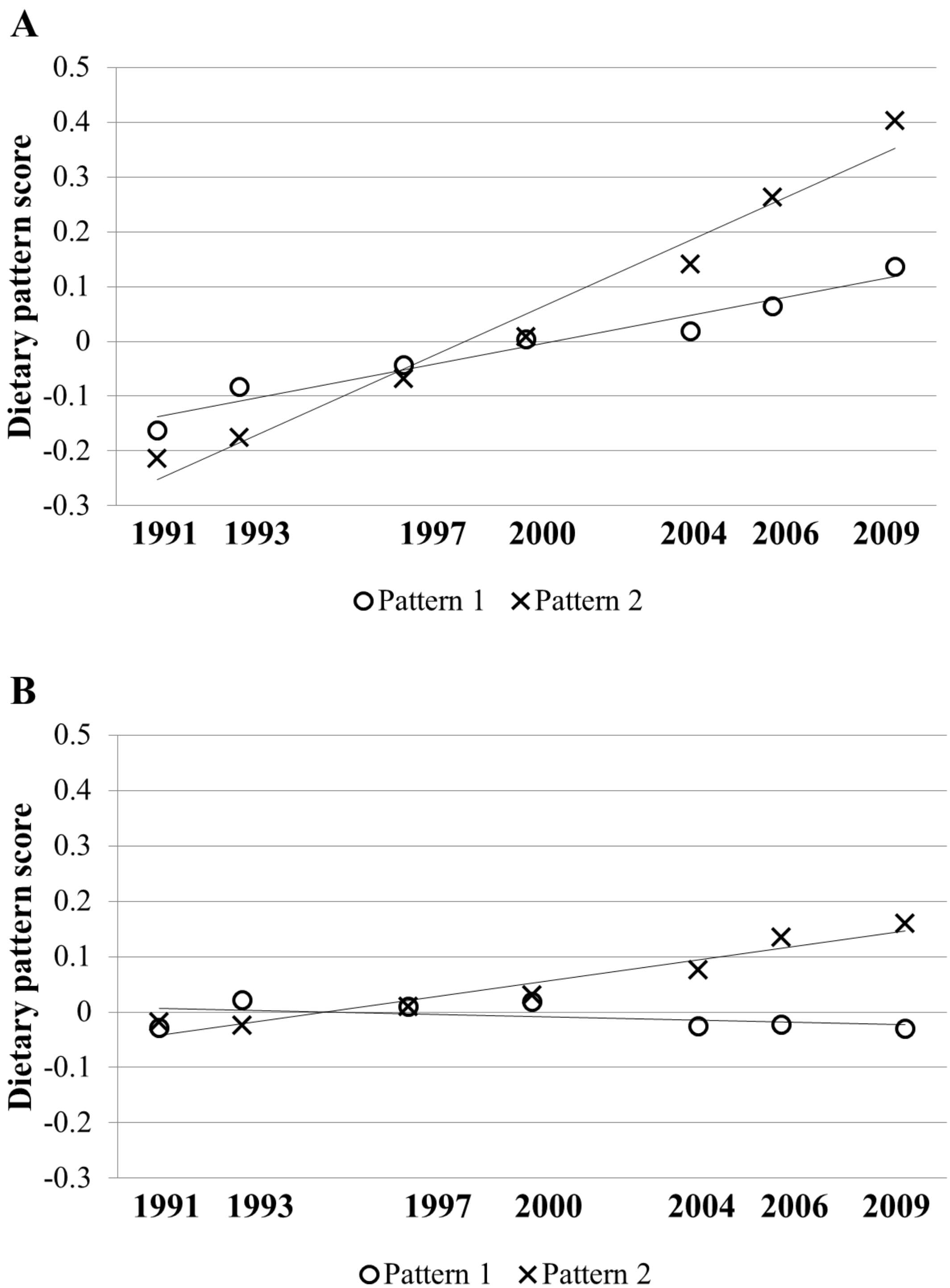
Supplemental Table 3.1 Examples of food items for each food group.

Food group	Examples of food items
Rice	White and brown rice
Wheat noodles	Wheat noodles
Wheat flour	Wheat flour
Wheat buns, breads	Bun, butter bread, salty bread
Cakes, cookies and pastries	Cookies, mooncake, fruit cake, chocolate cake, fruit pie
Deep-fried wheat	Deep-fried dough stick, deep-fried cake with red bean paste and sugar, deep-fried sweet sesame seed ball
Deep-fried rice and legumes	Deep-fried rice flour doughnut, deep-fried soybean, deep-fried broad bean
Corn and coarse grain	Corn, corn grits, corn flour, barley, oats, foxtail millet, sorghum
Starchy roots and tubers	Potato, yam, taro, lotus root, water chestnut, cassava, sweet potato
Fresh legumes	Soybean sprouts, peas with pod, mung bean sprouts
Dried legumes	Soybean flour, dried beans, beans flour, roasted broad bean
Legume products	Tofu, tofu products, red/mung bean paste
Nuts and seeds	Sesame, sunflower, watermelon seeds, lotus seeds, peanuts, walnuts, almonds, hazelnuts, pine-nuts, pistachios, cashew nuts
Starchy roots products and tubers products	Potato starch, lotus root starch, potato flour, corn starch, starch
Fresh vegetables, non-leafy	Cauliflower, tomatoes, cucumber, zucchini, mushrooms
Fresh vegetables, leafy	Spinach, ‘bok choy’, cabbage
Pickled, salted or canned vegetables	Canned tomato sauce, preserved vegetables, vegetables in soy sauce
Dried vegetables	Dried radish, dried bamboo shoot, dried lily
Seaweed	Fresh or dried seaweed
Fruits	Fresh and canned (no added sugar) fruits
Dried fruit	Dates, dried longan, raisins
Preserved fruit with added sugar	Dried and canned fruit with added sugar
Low-fat red meat	Low-fat beef, low-fat lamb, donkey, rabbit
High-fat red meat	High-fat beef, high-fat lamb
Low-fat pork	Pork tenderloin pork, pork tendons
High-fat pork	Pork belly, leg, rib chop
Organ meats	Liver, kidney, large intestine, blood
Processed meats	Sausages, ham, luncheon meat, dried meat, smoked meat
Poultry and game	Chicken, duck, goose
Eggs and eggs products	Whole eggs, yolk, white, preserved eggs
Fish and seafood	Fresh- and salt-water fish, dried fish, shellfish
Soy milk	Sweetened and un-sweetened soy milk
Animal-based milk	Cow milk, goat milk, skim milk, flavored milk
Dairy products	Cheese, yogurt

Sweetened dairy products	Ice cream
Western-style fast-food	Fried chicken, sandwich, hamburger, hotdog, pizza
Instant noodles and frozen dumplings	Instant noodles, frozen dumplings
Ready-to-eat cereals/porridge	Instant multigrain porridge, corn flakes, instant oatmeal
Salty snacks	Corn crisps, onion rings, potato chips,
Candy, sugar and other high-sugar foods	Jelly, jam, chocolate, honey, sugar, candies
Calorically-sweetened beverages	Fruit or flavored drinks, fruit juice, soft drinks
Low-caloric beverages	Tea, bottled water
Alcoholic beverages	Liquors, wine, vodka, cocktails, whiskey, beer

Figure 3.1 Mean factor scores over time.

Panel A: adjusted by geographical region and age in 1991. Panel B: additionally adjusted by number of food groups consumed.



Chapter 4. Using both Principal Component Analysis and Reduced Rank Regression to Study Dietary Patterns and Diabetes in Chinese Adults.

Overview

Diabetes is increasing in China and few studies have evaluated how dietary patterns relate to this disease in this population. We used the strengths of both Principal Component Analysis (PCA), to identify the eating patterns of the population; and Reduced Rank Regression (RRR), to derive a pattern that explained the variation in HbA1c, HOMA-IR and fasting glucose. We included 4,316 adults from the China Health and Nutrition Survey. Diet was measured over a 3-day period with 24-hr recalls and a household food inventory in 2006, and used to create dietary pattern scores using PCA and RRR. The outcomes were measured in 2009. The adjusted Odds Ratio for diabetes prevalence (HbA1c $\geq$6.5%), comparing the highest versus the lowest dietary pattern score quartile was 1.26 (0.76, 2.08) for a modern high-wheat pattern (from PCA, wheat products, fruits, eggs, milk and instant noodles/frozen dumpling), 0.76 (0.49, 1.17) for a traditional southern pattern (from PCA, rice, meat, poultry and fish) and 2.37 (1.56, 3.60) for the pattern derived with RRR pattern. By comparing the dietary pattern structure of RRR and PCA, we found that the RRR pattern was also behaviorally meaningful, it combined the deleterious effects of the modern high-wheat pattern (high intake of wheat buns & breads, deep-fried wheat, soy milk) with the deleterious effects of eating the opposite to the traditional southern (low intake of rice, poultry & game, fish & seafood). Our findings suggest that using both PCA and RRR provided useful insights when studying the association of dietary patterns with diabetes.

Introduction

Type 2 diabetes in China has been dramatically increasing over the last decades. Among adults, diabetes prevalence was 3%[69] in 1994 and 10% in 2007-08.[42] It is known that diet plays a key role in the prevention of diabetes, and attention has been devoted to understand which aspects of diet have the largest potential in helping prevent this disease.[70-72] The range of research goes from particular nutrients and food groups to dietary patterns.

The study of dietary patterns is appealing because they are closer approximations to the dietary exposures actually experienced by the population, as people do not consume single nutrients or foods but a combination of these in their daily diets. A common way to select the foods that will be represented in a given dietary pattern is by using statistical methods such as factor or principal component analysis (PCA). These methods derive linear functions of foods that best explain the variation in the intake of many foods or food groups, hence the resulting patterns represent foods that are typically consumed together by the population.[14,73] Although empirically derived dietary patterns are meaningful and describe the eating behavior of the population they might not be the most predictive of a disease. In contrast, Reduced Rank Regression (RRR) is similar to PCA in that is also a data reduction technique that determines linear functions of food groups, but the linear function of foods seeks to maximize the variation in outcome variables (e.g. disease related nutrients or biomarkers), and possibly explain only a moderate fraction of food groups variation.[74]

The strength of one method is the limitation of the other; PCA identifies patterns that have public health relevance because they describe the actual dietary patterns of the population whereas the foods in the RRR are not necessarily consumed together and hence could be behaviorally irrelevant. Conversely, the patterns from RRR are by definition associated with the outcome or response variables, which might not be the case for PCA patterns. Therefore both methods can complement each other and provide useful insights when compared side by side. For the PCA patterns, RRR patterns could help as a reference of the largest possible strength of association a data driven dietary pattern can have, which can help put in perspective the findings from PCA. On the contrary for the RRR, the PCA patterns can be a reference of which foods are eaten in combination, this can put in perspective how behaviorally meaningful the RRR dietary patterns are. In addition, RRR can help generate hypotheses about which key food components of a pattern are related to diabetes, and PCA can help identify whether these foods define the eating patterns of the population.

Few studies have looked at the association between dietary patterns and diabetes or insulin resistance among Chinese adults,[58,61,75] and none of these have used RRR. Therefore our aim in this study was to use both PCA and RRR and to complement each other by comparing the derived patterns and their strength of association with diabetes and insulin resistance. For the RRR we selected response variables that directly represent our outcomes of interest (HbA1c, fasting glucose and HOMA-IR). Typically RRR is used on intermediate response variables so that the dietary pattern can incorporate information about biological pathways.[73] However we were not interested in identifying biological pathways, but on identifying the dietary pattern that was most closely related to our outcome of interest. Therefore, our RRR dietary pattern should be considered more as an initial hypothesis and not as a pattern with established association.

Methods

Study design and participants

We included participants of the China Health and Nutrition Survey (CHNS), an ongoing study with detailed income, employment, education, demographic, health, and diet information elsewhere.[49] Briefly, the survey was designed to examine across space and time how economic and social changes are associated with a range of health behaviors. A multistage, random cluster process was used to draw the sample in 9 provinces. Survey protocols, instruments, and the process for obtaining informed consent for this study were approved by the institutional review committees of the University of North Carolina at Chapel Hill (UNC-CH) and the Chinese Institute of Nutrition and Food Safety (INFS), China Center for Disease Control and Prevention. Participants provided their written, informed consent.

Surveys were conducted in 1989, 1991, 1993, 1997, 2000, 2004, 2006 and 2009. Blood samples were collected for the first time in 2009, therefore in this analysis we only included 2006 and 2009 surveys. Our exposure of interest, dietary intake, and covariates were measured in 2006 and the diabetes related biomarkers were measured in 2009. Starting with 5,840 eligible subjects that were part of the survey in 2006 and 2009, 18-65 years old in 2006, not previously diagnosed with diabetes (n=194 were previously diagnosed) and not pregnant in 2006 or 2009 (n=40 were pregnant); we excluded those with missing dietary data (n=127) or biomarkers (n=743), not fasting before blood withdrawal (n=213), with missing BMI (n=273) or other covariates missing (n=167). Our final analytic sample was 4316. We considered ineligible those that in any wave (1991 to 2009) reported being previously diagnosed with diabetes or taking diabetes medications because treatment might have affected their dietary intake and/or biomarkers of glucose homeostasis.

Measurement of variables

Dietary assessment and food grouping

The dietary assessment in the CHNS is a combination of three consecutive days of 24-hour recalls at the individual level and a food inventory at household level performed over the same three day period. The three consecutive days were randomly allocated to start from Monday to Sunday. For the food inventory, all available foods at the household (purchased, stored or home produced) were measured on daily basis with digital scales. The changes in the household food inventory as well as the wastage were used to estimate total household food consumption. For the 24-hour recall, trained interviewers recorded the types, amounts, type of meal and place of consumption of all food items consumed. The amount of each dish was estimated from the household food inventory, based on the proportion of each dished the person reported to have consumed.

The food groups included in our analysis were based on a food grouping system developed specifically for the CHNS by researchers from UNC-CH and INFS[10] which separates foods into nutritional and behavioral meaningful food groups. For the dietary pattern analysis we did not include alcoholic beverages reported during three 24-hour recalls, because it was mostly consumed by males, with very low consumption for females. Nonetheless we included alcohol intake, reported from a frequency questionnaire as a covariate in our models.

Biomarkers

Blood samples were collected by venipuncture after an overnight fast. Glucose was measured on the serum with a glucose oxidase phenol 4-amnioantipyrine peroxidase kit (Randox, UK) in a Hitachi 7600 analyzer. HbA1c was measured on the whole blood by high-performance liquid chromatography with an automated glycohemoglobin analyzer (model HLC-723 G7; Tosoh, Tokyo, Japan). Insulin was measured on the serum by radioimmunology in a Gamma counter XH-6020 analyzer. The homeostasis model of insulin resistance was estimated [HOMA-IR=(fasting insulin (μU/ml)* fasting glucose (mmol/l))/ 22.5].[76]

We present results for HbA1c and not for glucose, and define diabetes based on HbA1c $\geq$6.5%.[77] Despite the controversies of using HbA1c as a diagnostic tool, it has the advantage over a single measure of glucose to capture long-term glycemic exposure.[78] In addition, HbA1c correlates well with the risk of long-term diabetes complications and it has been shown to be a reliable method for diabetes diagnosis in Chinese population.[79-81] Diabetes definition was only based on HbA1c because those on diabetes treatment or previously diagnosed were already excluded from the analyses.

Covariates

Physical activity was assessed with detailed self-report of time spent and intensity levels for occupational and domestic activities, and Metabolic Equivalents per week were estimated.[82] Income was assessed at the household level with a detailed questionnaire that included average salary for the last year for all household members, including bonus and other sources of income (i.e. home gardening). Level of urbanization was determined by an urbanicity scale that was developed for the CHNS, it includes components such as population density, economic activity, transportation infrastructure, sanitation, housing types, etc.[68] For all physical activity, income and urbanicity scale participants were classified into tertiles. BMI was estimated from measured weight and height; weight was measured without shoes in light clothing to the nearest 0.1 kg on a calibrated beam scale and height was measured also without shoes with a portable SECA stadiometer to the nearest 0.2 cm. Other demographic and lifestyle covariates included in the analysis were gender, age, geographical region (North, Central or South), education level, smoking status and alcohol intake.

Statistical analysis

Dietary patterns analysis was performed on 29 food groups. Most food groups had a high proportion of non-consumers; possibly due to the fact that diet intake was measured over a 3-day period. Therefore, for food groups with <80% of consumers we categorized food group intake into non-consumers vs. consumers and otherwise (rice, fresh non-leafy vegetables and fresh leafy vegetables) below vs. above the median. We divided the food groups for which the majority of the sample consumed as below or above median because there was little variation left in consumption vs. non-consumption. Food groups with ≤5% of consumers were not included in the dietary pattern analysis. For a full list of included and not included food groups, and their description see Supplemental Table 4.1.

PCA (*pca* command) was first performed in Stata 12.1 (StataCorp, College Station, TX) to evaluate the scree plot of the eigenvalues, based on this and interpretability we decided to retain two components. The analysis was repeated in the equivalent procedure PROC PLS with PCR method option (SAS 9.3, SAS Institute Inc., Cary, NC) in order to obtain the percent of variation explained in the food groups and in the response variables by each factor. RRR was performed with PROC PLS and RRR option with HbA1c, HOMA-IR and fasting glucose as response variables. Due to non-normality, all the response variables were first natural log-transformed. Because RRR can potentially derive patterns that are confounded by non-dietary factors,[83] we also used adjusted food groups with the residual method, a strategy previously used. [84,85] To estimate the residuals, we ran several logistic regressions with each binary food group as the dependent variable and geographical region, urbancity index, education and income as the independent variables. These residuals (difference between the observed and the predicted probability) were used as intake variables on the RRR procedure. Variable selection for the estimation of residuals was based in a previous work (Batis et al, unpublished data, 2013) where we found that in this population, these variables were the most influential on dietary patterns. We did not include energy intake in the residual method because we hypothesized it is in the causal pathway between the dietary pattern and diabetes. We adjusted by height, as an alternative to adjust by energy requirements but not by energy intake; because the pattern was unchanged with and without the inclusion of height, we present the pattern that did not include it. Also, because the RRR dietary pattern derived from the residuals was less confounded by non-dietary factors; on our main results we do not present the RRR performed with the original intake variables, but this is presented on Supplemental Tables 4.2 and 4.3. We did not use residuals in the PCA because here the aim is to describe how the population eats and there is no interest in adjusting by the factors that may influence these dietary patterns.

For RRR, the maximum number of factors we were able to obtain was three, corresponding to the number of response variables. We retained only the first factor because this one explained most of the variation in the response variables and was the only one with significant associations. Even if for PCA we retained two factors, the results are still comparable between the two methods, because the number of factors retained does not affect the structure of the derived patterns or the explained variation of each. For each individual a score was calculated for each dietary pattern (for each PCA and RRR) as a weighted sum of the food groups based on the factor loadings. The higher the score, the more closely the participant's diet conforms to the dietary pattern.

Although we knew the dietary pattern scores from RRR would had a stronger association with the outcomes, we ran multiple linear (for HbA1c and HOMA-IR) and logistic (for diabetes) regressions for each dietary pattern, to identify how different was the strength of association between RRR and PCA factors. Because HbA1c and HOMA-IR were natural log-transformed, the regression coefficients were multiplied by 100 so that they could be interpreted as the percent change in the outcome due to 1 unit increase in the independent variable. We present both by quartiles of the dietary pattern score and by the continuous increment in the score (1 standard deviation unit increase). First we adjusted by gender, smoking (yes/no), alcohol (≥3 times/week vs. <3 times/week), education (none, primary school, ≥ lower middle school), region (South, Central, North), age, income, urbancity index, physical activity (continuous). We did not adjust initially by BMI, because we hypothesized it was in the causal pathway, but to estimate the dietary pattern association independent of BMI, in a second model we additionally adjusted by BMI (continuous). The clustering at the household level was accounted for in the estimation of the variance.

Results

Compared to those with non-diabetic, those classified as diabetic were older, a higher proportion lived in the Central region, had a higher BMI, a higher proportion of males consumed alcohol regularly, had lower education, income, physical activity levels and lived in more urbanized areas (Table 4.1).

The first factor from PCA was inversely associated with the intake of rice and positively associated with wheat buns & breads, cakes/cookies and pastries, deep-fried wheat, fruits, eggs, soy milk, animal-based milk and instant noodles and frozen dumplings (Table 4.2). In this population we have previously found a similar pattern and named it “modern high-wheat” (Batis et al, unpublished data, 2013). The second factor from PCA was positively related to rice, high-fat pork, organ meats, poultry & game, fish & seafood and inversely associated with wheat flour, wheat buns & breads and corn & coarse grains. This pattern has been previously named “traditional southern” (Batis et al, unpublished data, 2013). The factor loadings from RRR seemed to be close to the modern high-wheat dietary pattern (PCA 1) and opposite to the traditional southern at the same time (PCA 2). As in the modern high-wheat, the RRR pattern was also inversely associated with rice and positively associated with wheat buns & breads, deep-fried wheat and soy milk; and in the opposite direction to the traditional southern it was negatively related to rice, poultry & game and fish & seafood. In addition, the RRR pattern was positively associated with wheat noodles and negatively associated with fresh legumes, items that were not related to the PCA patterns. Eggs & eggs products, was the only item that was associated in the opposite direction in RRR and the modern high-wheat pattern.

As expected the percent variation explained in food groups was higher for the PCA factors, and the percent variation explained in the responses tended to be higher for the RRR. The percent explained for HbA1c was higher in the PCA 2 than in the RRR, which could be related to using adjusted food groups in the RRR.

Both PCA factors (modern high-wheat and traditional southern) had a very strong association with HbA1c and Diabetes that was greatly weaken after adjustment by covariates, on the contrary the estimates for the RRR were only slightly closer to the null after adjustment, which is also related to use of residuals in RRR (Tables 4.3 and 4.4). For the adjusted estimates, comparing the fourth versus the first quartile, regression coefficients from the three dietary patterns were significantly different from zero for HbA1c; the association was positive for the modern high-wheat and RRR, and negative for the traditional southern dietary pattern. For HOMA-IR only the traditional southern had a negative association and the RRR pattern a positive one. For diabetes only the dietary pattern from RRR had a significant positive association. Comparing to the RRR dietary pattern, the strength of association of the PCA-based modern high-wheat pattern was 55%, 88% and 73% weaker for HbA1c, HOMA-IR and diabetes respectively; whereas for the traditional southern it was 41% and 68% weaker for HbA1c and diabetes and 13% stronger for HOMA-IR (for estimates from adjusted model 1, comparing 4^{th} vs 1th quartile). Additionally adjusting by BMI brought all the estimates closer to the null.

It is noteworthy that adjusting the food groups with the residual method changed the structure of derived pattern. In the RRR pattern on the original variables the factor loading for rice was -0.43, (which decrease to -0.22 when using residuals); and the only other food groups strongly associated with this dietary pattern were wheat flour, wheat buns/breads, deep-fried wheat, corn/coarse grain and soy milk. In addition, both the percent of variation explained in the food groups and response variables, and the strength of association with the outcomes for the unadjusted estimates was larger for the RRR on the original variables. However, adjusting by covariates greatly weaken the estimates to the point that they were similar between the two patterns, which showed that the RRR dietary pattern using the residuals was already less confounded (Supplemental Tables 4.2 and 4.3).

Discussion

In this study we used both PCA and RRR to study the association between dietary patterns and diabetes in China. With PCA we found two dietary patterns that relate to the eating behavior of this population; a modern high-wheat pattern was positively associated with HbA1c, whereas a traditional southern pattern was negatively associated with HbA1c and HOMA-IR. To put these associations in perspective, using the strength of association of the RRR as a reference, we found that the association of the PCA patterns was ~50% and ~70% weaker for HbA1c and diabetes respectively. However the negative association of the traditional southern pattern with HOMA-IR was comparable or even higher than the RRR one. On the other hand, using as a reference the structure of the PCA patterns as being the most behaviorally meaningful, we found that the RRR pattern was closely related to both PCA patterns in that it combines the deleterious effects of following the modern high-wheat pattern (high intake of wheat buns & breads, deep-fried wheat, soy milk) with the deleterious effects of following a diet opposite to the traditional southern (low intake of rice, poultry & game, fish & seafood). This gives public health relevance to the RRR pattern because it was not only associated with markers of diabetes but it was also related to dietary patterns actually followed by this population.

Another useful aspect of comparing the results from PCA and RRR is to identify which are the food groups that differ between them. Based on this we can generate the hypothesis that reduced fresh legumes and increased wheat noodles, even if not important characteristics of behavioral dietary pattern they are important part of dietary pattern associated with diabetes. On the other hand, cakes, cookies & pastries, fruits, animal-based milk, instant noodles & frozen dumplings, high-fat pork and organ meats even if important for defining behavioral dietary patterns, they are not a key part of a diabetes-related dietary pattern in this population. The reasons behind these findings could be both related to the intrinsic nutritional characteristics of these food groups or to their distribution of consumption in this population. For example cakes, cookies & pastries and instant noodles & frozen dumplings are items that could potentially be related to diabetes, but the proportion of consumers for these items was 8% and 12% respectively, so perhaps there was not enough variability in their consumption to explain the variation in the outcomes.

With RRR we found that a dietary pattern high in wheat products and low in legumes, poultry and fish was positively associated with diabetes, this is consistent with the literature. Evidence suggests that glycemic index and staples like rice, noodles and steamed bread and bread are associated with greater risk of diabetes, whereas higher intake of dietary fiber and legumes have a protective effect against the disease.[86-90] In our analysis, the RRR patterns associated with diabetes markers were positively associated with many refined carbohydrates-based foods such as wheat noodles, wheat buns & breads and deep-fried wheat and negatively associated with fresh legumes. Fish & seafood were inversely related to our RRR pattern, in meta-analyses that stratified by region it has been reported that fish intake has a protective effect against diabetes in Asian countries.[60,91,92] Poultry was also inversely related to our RRR pattern, and there is evidence in Chinese population that the intake of poultry is associated with decreased risk of type 2 diabetes.[93]

On the other hand, some of the food groups that were related to our RRR pattern are inconsistent with what previous studies have reported. Eggs intake, particularly when comparing the intake of ≥1 per day versus <2 per week was associated with higher odds of diabetes prevalence in a cross-sectional study in the Jiangsu province in China.[94] In our analysis, eggs intake were inversely related to the RRR pattern, it is possible that comparing any consumption versus no consumption at all, as in our case, yields a different association; or that subjects that consume eggs are less likely to consume red meat. In Shanghai, soy milk has been found to be protective for diabetes incidence;[87] however in our analysis this food was positively associated with the RRR pattern. One reason could be that sweetened soy milk is gaining popularity in China, however in our sample only 8% of the soy milk consumed was sweetened. Another possibility is that in China it is common to have deep fried dough accompanied with soy milk for breakfast, and this could be the reason why in our dietary pattern analysis both food groups remained together in the same pattern.

Moreover, in our analysis, rice was inversely associated with the RRR patterns. Yet, rice intake has been associated with higher risk of diabetes in the US, China and Japan.[95] In a study in Shanghai rice was the top contributor to glycemic load in the diet.[86] However a randomized trial substituting white rice with brown rice had no effect on metabolic risk factors.[96] In addition, studies in China assessing dietary patterns have found that subjects following a "green water" pattern characterized by a high intake of rice and vegetables, moderate intake of fish, poultry and pork and low intake of wheat and other cereals had the lowest prevalence of glucose tolerance abnormalities in the 2002 China National Nutrition and Health Survey[75]. Furthermore a pattern inversely related with rice and high in wheat was positively associated with insulin resistance in the Jiangsu province.[61] So it is possible that rice evaluated independently is associated with increased risk, but that its intake is related to a more healthful dietary pattern.

A drawback from RRR is that the derived patterns have the potential to be confounded by other non-dietary factors. For example, in our case it is also possible that rice was inversely associated with the RRR pattern partially due to the fact that in the South the consumption of rice is high and the prevalence of diabetes is low. The concept of using residuals is to first adjust the food group before using them to derive the dietary patterns. The impact of using the residuals can vary by population or research question, if confounding factors are not strongly related to both the food groups and the response variables then using or not the residuals should not make a difference, as previous studies have found.[84] In our analysis, it made a difference to use the residual method to adjust the food groups before performing the RRR method. The structure of the dietary pattern differed (i.e. the loading for rice became weaker and other food groups emerged), and the change in estimate from the unadjusted to the adjusted by covariates was much larger for the RRR performed on the original intake variables than for the RRR from the residuals. Even when the adjusted estimates were relatively similar, it is preferable to have a dietary pattern that is already less biased, and therefore using the residuals was a useful approach.

Four studies have used RRR on biomarkers and dietary intake data to derive dietary patterns that predict diabetes in American and European populations.[84,85,97,98] They used diet intake and response variables that were measured simultaneously at baseline among individuals not diagnosed with type 2 diabetes, then the relationship between the dietary pattern score and incident diabetes during several years of follow-up was assessed. Some of the response variables used were inflammatory biomarkers,[97,98] HDL cholesterol,[85,98] HbA1c, adiponectin,[98] HOMA-IR,[84] BMI, fasting glucose, triglycerides and blood pressure.[85] All of these studies found that the RRR pattern was predictive of incident diabetes. Common aspects in their dietary patterns were that all of them were associated with refined grains and caloric soft drinks, 3 of them were associated with processed meat, and half of them were associated with red meat, low-caloric soft drinks, and negatively associated with vegetables and wine. However even when the patterns had items in common, Imamura et al. found that the patterns from the studies conducted in Europe[84,98] were not generalizable to their US population.[85] Therefore, the pattern that we found in a Chinese population might be even less comparable; still we also found that refined carbohydrates were a very important part of this dietary pattern.

Several studies have compared dietary patterns derived from RRR and PCA with different health outcomes; they have mainly focused on comparing which method yields more significant associations.[74,99-103] All except for one[99] concluded that RRR derived stronger or more statistical significant patterns. Still when comparing only the first factor from PCA and RRR the estimates in all these studies were always stronger with RRR. More importantly, from the studies that present the factor loadings for both methods, in the majority[99,101,102] their first RRR and PCA factors were relatively similar. In our analysis, we also found that the RRR patterns were closely related to the PCA patterns, this is an important finding because it means that patterns from RRR are also behaviorally meaningful. Therefore when using RRR, it should be encouraged to compare it with PCA patterns and confirm if the RRR pattern has a behavioral significance in the population under study.

To best of our knowledge this is the first study that compares RRR and PCA in relation to diabetes in Chinese population. A strength of our analysis is that we were not limited to a specific urban area or province; the surveyed provinces in our sample represent 56% of the Chinese population. Another strength is the longitudinal nature of the study, the temporal sequence is unambiguous because diet was measured in 2006 and the outcome was measured in 2009. Although this design has advantages over cross-sectional studies, a key limitation of our analysis is that biomarkers of glucose homeostasis were only measured for the first time in 2009; therefore we could not distinguish between incident and prevalent diabetes. Nonetheless, particularly when assessing HbA1c and HOMA-IR, even if subjects had HbA1c ≥6.5% before 2009 and were already undiagnosed diabetics, diet could still influence their biomarker values. In addition, to avoid reverse causality (i.e. subjects improving their diet because of diabetes diagnosis) we excluded all the subjects that reported being previously diagnosed with diabetes.

The dietary assessment in the CHNS is very detailed and precise, however because it covers only 3 days of intake it is not the best measure of usual intake.[10,49,104] However we found that the variation explained in the response variables and food groups was comparable to what other studies have found using Food Frequency Questionnaire.[74,84] Another limitation of not having usual intake, is that proportion of consumers during 3 days was very low (≤ 5%) for many key food groups such as preserved fruit with added sugar, processed meats, sweetened dairy products, Western-style fast-food, salty snacks, ready-to-eat cereals/porridge, calorically-sweetened beverages, and low-caloric beverages. It is possible that these foods are important for the development of diabetes in this population, however because over a 3-day period they were not widely consumed we were not able to include these in our dietary pattern analysis. Particularly important are caloric beverages, although compared to the US the intake is still low in China, it was increasing between 2000 and 2010.[105] If this trend continues future research might find that caloric beverages are related to diabetes in China, as studies among Chinese Singaporeans have already reported.[106]

In sum, we found that using both methods gave important insights. The aim of each PCA and RRR is different and their results complement each other. We found that in comparison to the RRR, the method that aims to derive patterns that are related to the outcome, a traditional southern dietary pattern derived with PCA had a comparable strength of association with insulin resistance (negative association). Also, that in comparison with the PCA, the method that aims to describe the eating behavior of the population, the foods identified in the RRR were in line with the type of foods the population eats in combination. According to our findings, we can hypothesize that from the modern high-wheat pattern the key combination of foods associated with diabetes is wheat buns & breads, deep-fried wheat and soy milk; and from the traditional southern pattern, the key protective combination of foods is rice, poultry and fish. Because these sets of foods groups are typically consumed together in the population it could be possible to identify the subjects at higher risk of following this pattern and intervene accordingly if further evidence warrants it.

Tables

Table 4.1 Baseline Characteristics of Participants by Diabetes Status

	Diabetes (HbA1c ≥6.5%)	
	No (n= 4,071)	**Yes (n=245)**
Age (years), mean ± SD	46.5 ± 10.5	51.7 ± 8.6
Region, %		
South	45.9	22.0
Central	32.0	57.6
North	22.1	20.4
Male, %	45.3	49.8
BMI (kg/m^2), mean ± SD	23.2 ± 3.1	25.7 ± 3.8
Highest level of education attained, %		
None	20.5	25.3
Primary school	20.5	25.3
≥Lower middle school	59.0	49.4
Income, %		
Low	33.33	37.55
Medium	33.6	30.2
High	33.06	32.24
Urbancity, %		
Low	33.8	26.1
Medium	33.7	34.7
High	32.6	39.2
Currently smoking, %		
Female	3.0	3.3
Male	58.0	58.2
Alcohol intake ≥3 times/week, %		
Female	2.0	1.6
Male	30.9	37.7
Physical Activity, %		
Low	32.7	40.8
Medium	33.4	32.7
High	34.0	26.5

Table 4.2. Factor loadings[a] and explained variation of dietary patterns obtained from Principal Component Analysis and Reduced Rank Regression.

	Principal Component Analysis		Reduced Rank Regression[b]
	Modern high-wheat	Traditional southern	
Food groups			
Rice	-0.25	0.34	-0.22
Wheat noodles	—	—	0.30
Wheat flour	—	-0.36	—
Wheat buns, breads	0.33	-0.26	0.46
Cakes, cookies and pastries	0.28	—	—
Deep-fried wheat	0.38	—	0.22
Corn and coarse grain	—	-0.30	—
Fresh legumes	—	—	-0.24
Fruits	0.32	—	—
High-fat pork	—	0.26	—
Organ meats	—	0.25	—
Poultry and game	—	0.31	-0.37
Eggs and eggs products	0.25	—	-0.23
Fish and seafood	—	0.37	-0.29
Soy milk	0.34	—	0.24
Animal-based milk	0.26	—	—
Instant noodles and frozen dumplings	0.23	—	—
Explained variation in food groups, %	8.47	7.64	4.42
Explained variation in responses,%			
HbA1c	0.96	2.95	1.40
HOMA-IR	0.08	0.09	0.41
Fasting glucose	0.26	0.18	0.62

[a]Factor loadings < |0.20| not shown. The following food groups had factor loadings <|0.20| in all patterns and are not shown in the table: Starchy roots and tubers, Dried legumes, Legume products, Nuts and seeds, Starchy roots products and tubers products, Fresh vegetables, non-leafy, Fresh vegetables, leafy, Pickled, salted or canned vegetables, Dried vegetables, High-fat red meat, Low-fat pork, Processed meats.

[b]Performed on residuals estimated for each food group with a multiple regression including geographical region, urbanicity, income and education.

Table 4.3 Percent change in HbA1c and HOMA-IR related to quartiles of dietary pattern score and linear dietary pattern score increase (1 SD)

	quartile 1	quartile 2	quartile 3	quartile 4	Dietary pattern score (1 SD increase)
			HbA_{1c}, % change (95% Confidence Interval)[a]		
PCA, Modern high-wheat					
Unadjusted model	0	2.00 (0.94, 3.05)	3.32 (2.26, 4.38)	4.29 (3.26, 5.32)	1.32 (0.97, 1.68)
Adjusted model 1	0	0.76 (-0.25, 1.77)	1.50 (0.39, 2.61)	1.68 (0.51, 2.86)	0.32 (-0.09, 0.74)
Adjusted model 2	0	0.53 (-0.45, 1.52)	1.27 (0.19, 2.36)	1.43 (0.28, 2.57)	0.30 (-0.10, 0.71)
PCA, Traditional southern					
Unadjusted model	0	-3.88 (-4.98, -2.78)	-5.96 (-6.99, -4.93)	-5.91 (-6.99, -4.82)	-2.20 (-2.60, -1.79)
Adjusted model 1	0	-1.56 (-2.67, -0.45)	-2.32 (-3.45, -1.19)	-2.21 (-3.40, -1.02)	-0.79 (-1.25, -0.33)
Adjusted model 2	0	-1.35 (-2.44, -0.26)	-1.96 (-3.07, -0.84)	-1.88 (-3.06, -0.71)	-0.72 (-1.18, -0.26)
RRR[b]					
Unadjusted model	0	1.74 (0.71, 2.77)	3.29 (2.30, 4.29)	4.04 (3.03, 5.05)	1.45 (1.09, 1.82)
Adjusted model 1	0	1.95 (1.02, 2.88)	3.22 (2.32, 4.13)	3.75 (2.84, 4.66)	1.45 (1.13, 1.77)
Adjusted model 2	0	1.75 (0.84, 2.66)	2.97 (2.09, 3.85)	3.44 (2.56, 4.31)	1.32 (1.01, 1.64)
			HOMA-IR, % change (95% Confidence Interval)[a]		
PCA, Modern high-wheat					
Unadjusted model	0	7.16 (0.66, 13.66)	15.50 (8.94, 22.06)	18.69 (12.10, 25.29)	6.05 (3.79, 8.32)
Adjusted model 1	0	1.79 (-4.55, 8.13)	5.80 (-0.75, 12.34)	1.13 (-6.28, 8.53)	-1.06 (-3.75, 1.63)
Adjusted model 2	0	-0.36 (-6.53, 5.80)	3.66 (-2.65, 9.97)	-1.28 (-8.40, 5.84)	-1.24 (-3.85, 1.37)
PCA, Traditional southern					
Unadjusted model	0	-5.16 (-11.39, 1.07)	-11.52 (-17.93, -5.11)	-5.35 (-11.35, 0.65)	-1.95 (-4.08, 0.17)
Adjusted model 1	0	-4.01 (-10.57, 2.55)	-12.64 (-19.99, -5.29)	-10.92 (-18.54, -3.30)	-4.18 (-6.95, -1.42)
Adjusted model 2	0	-1.99 (-8.20, 4.21)	-9.21 (-16.30, -2.12)	-7.87 (-15.25, -0.49)	-3.56 (-6.25, -0.87)
RRR[b]					

Unadjusted model	0	2.68 (-3.30, 8.65)	10.93 (4.68, 17.18)	10.81 (4.36, 17.25)	4.85 (2.57, 7.14)
Adjusted model 1	0	5.95 (0.16, 11.73)	13.08 (7.08, 19.08)	9.67 (3.47, 15.87)	4.76 (2.58, 6.94)
Adjusted model 2	0	4.04 (-1.54, 9.63)	10.65 (4.85, 16.45)	6.66 (0.63, 12.68)	3.56 (1.44, 5.68)

[a]Regression performed with logarithms of HbA1c and HOMA-IR, therefore coefficients are interpreted as % change

[b]Performed on residuals estimated for each food group with a multiple regression including geographical region, urbanicity, income and education.

Model 1: adjusted by gender, smoking, alcohol, education, region, age, income, urbancity index, physical activity; Model 2: adjusted by variables in Model 1 plus BMI

Table 4.4 Association between Diabetes (HbA1c ≥6.5%) and quartiles of dietary pattern score and linear dietary pattern score increase (1 SD)

	quartile 1	quartile 2	quartile 3	quartile 4	Dietary pattern score (1 SD increase)
			Diabetes prevalence (%)		
PCA, Modern high-wheat	3.8	4.9	6.6	7.4	
PCA, Traditional southern	8.4	5.8	4.1	4.4	
RRR[b]	3.2	5.5	6.7	7.4	
			Odds Ratio (95% Confidence Interval)		
PCA, Modern high-wheat					
Unadjusted model	1	1.31 (0.86, 1.98)	1.79 (1.20, 2.65)	2.03 (1.38, 2.98)	1.20 (1.06, 1.35)
Adjusted model 1	1	1.06 (0.67, 1.68)	1.30 (0.82, 2.07)	1.26 (0.76, 2.08)	1.01 (0.86, 1.18)
Adjusted model 2	1	1.00 (0.63, 1.59)	1.22 (0.77, 1.95)	1.13 (0.68, 1.86)	0.99 (0.84, 1.16)
PCA, Traditional southern					
Unadjusted model	1	0.67 (0.48, 0.94)	0.46 (0.32, 0.67)	0.50 (0.34, 0.71)	0.74 (0.65, 0.85)
Adjusted model 1	1	0.90 (0.62, 1.32)	0.74 (0.48, 1.15)	0.76 (0.49, 1.17)	0.88 (0.75, 1.05)
Adjusted model 2	1	0.93 (0.64, 1.35)	0.82 (0.52, 1.29)	0.86 (0.54, 1.35)	0.91 (0.76, 1.09)
RRR[b]					
Unadjusted model	1	1.78 (1.16, 2.73)	2.20 (1.45, 3.33)	2.46 (1.63, 3.71)	1.36 (1.19, 1.55)
Adjusted model 1	1	1.86 (1.20, 2.88)	2.21 (1.44, 3.39)	2.37 (1.56, 3.60)	1.35 (1.19, 1.53)
Adjusted model 2	1	1.74 (1.12, 2.71)	2.08 (1.35, 3.21)	2.15 (1.41, 3.29)	1.31 (1.15, 1.49)

[b]Performed on residuals estimated for each food group with a multiple regression including geographical region, urbanicity, income and education.

Model 1: adjusted by gender, smoking, alcohol, education, region, age, income, urbancity index, physical activity; Model 2: adjusted by variables in Model 1 plus BMI

Supplemental Table 4.1 Examples of food items for each food group included and not included in dietary pattern analysis.

	Examples of food items
Food groups included in dietary pattern analysis	
Rice	White and brown rice
Wheat noodles	Wheat noodles
Wheat flour	Wheat flour
Wheat buns, breads	Bun, butter bread, salty bread
Cakes, cookies and pastries	Cookies, mooncake, fruit cake, chocolate cake, fruit pie
Deep-fried wheat	Deep-fried dough stick, deep-fried cake with red bean paste and sugar, deep-fried sweet sesame seed ball
Corn and coarse grain	Corn, corn grits, corn flour, barley, oats, foxtail millet, sorghum
Starchy roots and tubers	Potato, yam, taro, lotus root, water chestnut, cassava, sweet potato
Fresh legumes	Soybean sprouts, peas with pod, mung bean sprouts
Dried legumes	Soybean flour, dried beans, beans flour, roasted broad bean
Legume products	Tofu, tofu products, red/mung bean paste
Nuts and seeds	Sesame, sunflower, watermelon seeds, lotus seeds, peanuts, walnuts, almonds, hazelnuts, pine-nuts, pistachios, cashew nuts
Starchy roots products and tubers products	Potato starch, lotus root starch, potato flour, corn starch, starch
Fresh vegetables, non-leafy	Cauliflower, tomatoes, cucumber, zucchini, mushrooms
Fresh vegetables, leafy	Spinach, ‘bok choy’, cabbage
Pickled, salted or canned vegetables	Canned tomato sauce, preserved vegetables, vegetables in soy sauce
Dried vegetables	Dried radish, dried bamboo shoot, dried lily
Fruits	Fresh and canned (no added sugar) fruits
Low-fat red meat	Low-fat beef, low-fat lamb, donkey, rabbit
High-fat red meat	High-fat beef, high-fat lamb
Low-fat pork	Pork tenderloin pork, pork tendons
High-fat pork	Pork belly, leg, rib chop
Organ meats	Liver, kidney, large intestine, blood
Poultry and game	Chicken, duck, goose
Eggs and eggs products	Whole eggs, yolk, white, preserved eggs
Fish and seafood	Fresh- and salt-water fish, dried fish, shellfish
Soy milk	Sweetened and un-sweetened soy milk
Animal-based milk	Cow milk, goat milk, skim milk, flavored milk
Instant noodles and frozen dumplings	Instant noodles, frozen dumplings
Food groups not included in dietary pattern analysis*	
Deep-fried rice and legumes	Deep-fried rice flour doughnut, deep-fried soybean, deep-fried

	broad bean
Dried fruit	Dates, dried longan, raisins
Preserved fruit with added sugar	Dried and canned fruit with added sugar
Seaweed	Fresh or dried seaweed
Processed meats	Sausages, ham, luncheon meat, dried meat, smoked meat
Dairy products	Cheese, yogurt
Sweetened dairy products	Ice cream
Western-style fast-food	Fried chicken, sandwich, hamburger, hotdog, pizza
Salty snacks	Corn crisps, onion rings, potato chips,
Ready-to-eat cereals/porridge	Instant multigrain porridge, corn flakes, instant oatmeal
Calorically-sweetened beverages	Fruit or flavored drinks, fruit juice, soft drinks
Low-caloric beverages	Tea, bottled water
Candy, sugar and other high-sugar foods	Jelly, jam, chocolate, honey, sugar, candies
Alcoholic beverages	Liquors, wine, vodka, cocktails, whiskey, beer

*Not included because they had ≤5% of consumers, except for alcoholic beverages that was not included because it was mostly consumed by males.

Supplemental Table 4.2 Factor loadings and explained variation of dietary pattern obtained from Reduced Rank Regression performed on the original intake variables

Food groups	Factor loadings[a]
Rice	-0.43
Wheat noodles	–
Wheat flour	0.43
Wheat buns, breads	0.46
Cakes, cookies and pastries	–
Deep-fried wheat	0.29
Corn and coarse grain	0.35
Fresh legumes	–
Fruits	–
High-fat pork	–
Organ meats	–
Poultry and game	–
Eggs and eggs products	–
Fish and seafood	–
Soy milk	0.21
Animal-based milk	–
Instant noodles and frozen dumplings	–
Explained variation in food groups, %	6.38
Explained variation in responses,%	
HbA1c	5.45
HOMA-IR	0.29
Fasting glucose	1.04

[a]Factor loadings < |0.20| not shown. The following food groups had factor loadings <|0.20| in all patterns and are not shown in the table: Starchy roots and tubers, Dried legumes, Legume products, Nuts and seeds, Starchy roots products and tubers products, Fresh vegetables, non-leafy, Fresh vegetables, leafy, Pickled, salted or canned vegetables, Dried vegetables, High-fat red meat, Low-fat pork, Processed meats.

Supplemental Table 4. 3 Association between dietary pattern score from Reduced Rank Regression performed on the original intake variables and HbA1c, HOMA-IR and Diabetes (HbA1c ≥6.5%).

	quartile 1	quartile 2	quartile 3	quartile 4	continuous
		HbA_{1c}, % change (95% Confidence Interval)[a]			
Unadjusted model	0	1.98 (0.99, 2.97)	4.35 (3.32, 5.38)	7.77 (6.72, 8.82)	3.01 (2.64, 3.39)
Adjusted model 1	0	1.50 (0.56, 2.44)	2.37 (1.38, 3.36)	3.52 (2.46, 4.59)	1.44 (1.06, 1.83)
Adjusted model 2	0	1.43 (0.51, 2.35)	2.28 (1.32, 3.25)	3.15 (2.11, 4.19)	1.30 (0.93, 1.68)
		HOMA-IR, % change (95% Confidence Interval)a			
Unadjusted model	0	3.57 (-2.74, 9.88)	8.83 (2.58, 15.07)	18.52 (12.11, 24.92)	7.47 (5.28, 9.66)
Adjusted model 1	0	3.10 (-3.05, 9.26)	5.30 (-1.18, 11.78)	11.96 (4.43, 19.50)	4.83 (2.20, 7.46)
Adjusted model 2	0	2.42 (-3.53, 8.36)	4.47 (-1.78, 10.73)	8.45 (1.18, 15.72)	3.49 (0.94, 6.04)
		Diabetes prevalence (%)			
	3.0	4.4	6.4	9.0	
		Odds Ratio (95% Confidence Interval)			
Unadjusted model	1	1.49 (0.94, 2.35)	2.23 (1.46, 3.43)	3.24 (2.15, 4.87)	1.56 (1.38, 1.77)
Adjusted model 1	1	1.35 (0.83, 2.19)	1.60 (1.00, 2.55)	1.74 (1.09, 2.78)	1.25 (1.08, 1.44)
Adjusted model 2	1	1.35 (0.83, 2.20)	1.54 (0.96, 2.47)	1.54 (0.96, 2.48)	1.20 (1.03, 1.39)

[a]Regression performed with logarithms of HbA1c and HOMA-IR, therefore coefficents are interpreted as % change

Model 1: adjusted by gender, smoking, alcohol, education, region, age, income, urbancity index, physical activity; Model 2: adjusted by variables in Model 1 plus BMI

Chapter 5. Dietary Pattern Trajectories during 15 years of Follow-up and HbA1c, Insulin Resistance and Diabetes Prevalence among Chinese Adults.

Overview

We used an innovative technique to capture trajectories of a dietary pattern and assessed their relationship to HbA1c, insulin resistance and diabetes prevalence. We included 4096 adults with 3 to 6 waves of diet data (1991 to 2006) and biomarkers measured in 2009 from the China Health and Nutrition Survey. Diet was assessed by a household food inventory and three 24-h recalls. A dietary pattern in 2006 that positively predicted diabetes biomarkers was derived with Reduced Rank Regression, which yielded a pattern high in wheat products and soy-milk; and low in rice, legumes, poultry, eggs and fish. A score for this pattern was estimated for each subject at each wave and with Latent Class Trajectory Analysis subjects with similar diet pattern score trajectories over time were grouped in 5 classes. Three classes had parallel stable scores over time and two had a positive slope. Among two classes with similar scores in 2006, the one with lower initial scores had significantly lower HbA1c [-1.64 (-3.17, -0.11)], and non-significantly lower HOMA-IR [-6.47 (-17.37, 4.42)] and odds of diabetes [0.86 (0.44, 1.67)]. The results suggest that the different dietary trajectories that subjects followed are associated with HbA1c, even when their trajectories have the same end point.

Introduction

China is experiencing a rapid increase in diabetes, and for a country of over 1.3 billion people, the disease affects a significant number of persons. By 2007-2008, it was estimated that 92 million Chinese adults were diabetic.[107,108] In addition, compared to Western populations, Asians have higher total body fat, visceral fat and insulin resistance at lower BMIs.[6,7] Therefore, research focusing on a more comprehensive understanding on how modifiable risk factors, such as diet, affect diabetes is needed in this country.

Studying diet-disease associations with dietary patterns, instead of single nutrients or food groups, is appealing because the effect of a single nutrient/food may be too small to detect. In addition, because individuals consume a wide variety of foods rather than single nutrients or foods, dietary patterns resemble more closely the actual eating behavior.[12-14,109-111] However, most of the studies associating dietary patterns with health outcomes include diet from only point in time[12]; hence they are not able to capture how individuals change or maintain their dietary pattern over time and how this affects their health risks.

Only few studies have assessed the health effects of inter-individual changes over time in dietary patterns. One study evaluated the effect of transitioning between different clusters of dietary patterns.[112] Others assessed the effect of each individual slope and intercept or the difference between two time points in their dietary patterns' scores.[20,30,113] In the present study, we explored the application of latent class trajectory analysis (LCTA). This method simultaneously fits trajectory curves for every person, and identifies subgroups of subjects that share similar trajectories.[35] This technique has been used previously to characterize trajectories of BMI[37,114] or other social characteristics such as alcoholism,[38] depression,[39] delinquency,[40] etc, but to the best of our knowledge it has not been used for dietary patterns before.

We used 6 repeated measures of diet over 15 years of follow-up (1991 to 2006) to identify groups of individuals with similar trajectories of their dietary pattern's score over time, and then we estimated the association of each trajectory class with HbA1c, insulin resistance (HOMA-IR) and diabetes prevalence in 2009. To be able to evaluate whether the long-term trajectories of a dietary pattern were associated with our outcomes, we needed in the first place a dietary pattern with established association to diabetes in this population. Therefore we used a dietary pattern previously derived from Reduced Rank Regression (RRR) using as response variables HbA1c, HOMA-IR and glucose from 2009 and dietary intake in 2006 (Batis et al, unpublished data, 2013). Because RRR identifies linear combinations of foods that explain the variation in the response variables, as expected this dietary pattern in 2006 was associated with HbA1c, insulin resistance (HOMA-IR) and diabetes prevalence. Our motivation was that individuals may have a similar dietary pattern scores at one point in time (i.e. 2006), but some of them could have recently adopted this pattern and some could have followed the same pattern for years. It is important to understand if the associated health risks for these two types of trajectories are different.

Methods

Study design and participants

The China Health and Nutrition Survey (CHNS) is an ongoing longitudinal study. A multistage, random cluster process was used to draw the sample in 9 provinces. Survey protocols, instruments, and the process for obtaining informed consent were approved by the institutional review committees of the University of North Carolina at Chapel Hill (UNC-CH) and the Chinese Institute of Nutrition and Food Safety (INFS), China Center for Disease Control and Prevention. Additional details about the CHNS data are provided elsewhere.[49]

Surveys were conducted in 1991, 1993, 1997, 2000, 2004, 2006 and 2009. All waves had identical clinical, dietary and anthropometric data from each household member. Blood samples were collected for the first time in 2009, hence our main exposure, dietary intake, and covariates were measured from 1991 to 2006 and our outcomes the biomarkers of glucose homeostasis were measured in 2009. We included all subjects that had at least 3 waves of dietary data from 1991 to 2006; had complete biomarkers in 2009 and were fasting before blood collection; were 18-65 years old in all of the waves they were included from 1991 to 2006, were not previously diagnosed with diabetes, and were not pregnant in 2009. From the 4,096 subjects in our final sample, 40% had all the 6 waves of diet complete and 17% had only three waves complete. We did not include those that reported being previously diagnosed with diabetes in 2009 or before.

Dietary assessment and food grouping

A combination of three consecutive 24-hour recalls at the individual level and a food inventory at the household level were performed over the same three day period, which was randomly allocated to start from Monday to Sunday. For the food inventory, all available foods in the household were measured on daily basis, and wastage was considered to estimate the total consumption. For the 24-hour recall, trained interviewers recorded the types, amounts, type of meal and place of consumption of all food items consumed. For dishes prepared at home the amount for each person was estimated based on what they report and the household food inventory.

The food groups included in our analysis were based on a food grouping system developed specifically for the CHNS by researchers from UNC-CH and INFS,[10] this system separates foods into nutritional and behavioral meaningful food groups. Alcoholic beverages were not included in the dietary pattern, because it was mostly consumed by males, with very low consumption for females. Nonetheless we included alcohol intake, reported from a frequency questionnaire as a covariate in our models.

Biomarkers

Blood samples were collected by venipuncture after an overnight fast. Glucose was measured on the serum with a glucose oxidase phenol 4-amnioantipyrine peroxidase kit (Randox, UK) in a Hitachi 7600 analyzer. HbA1c was measured on the whole blood by high-performance liquid chromatography with an automated glycohemoglobin analyzer (model HLC-723 G7; Tosoh, Tokyo, Japan). And insulin was measured on the serum by radioimmunology in a Gamma counter XH-6020 analyzer. The homeostasis model of insulin resistance was estimated [HOMA-IR=(fasting insulin (μU/ml)* fasting glucose (mmol/l))/ 22.5].[76] We present results for HbA1c and not for glucose, and define diabetes based on HbA1c ≥6.5%.[77] HbA1c is more suitable than a single measure of glucose because it captures long-term glycemic exposure. In addition HbA1c correlates well with the risk of long-term diabetes complications and it has been shown to be a reliable method for diabetes diagnosis in Chinese population.[79-81]

Covariates

Demographic and lifestyle covariates included in the analysis were gender, age, geographical region (North, Central or South), education level, smoking status and alcohol intake. For physical activity detailed time spent and intensity levels were collected for occupational and domestic activities, and Metabolic Equivalents per week were estimated.[82] Income was assessed at the household level with a detailed questionnaire that included average salary for the last year. Level of urbanization was determined by an urbanicity index that was developed for the CHNS, it includes components such as population density, economic activity, transportation infrastructure, sanitation, housing types, etc.[68] BMI was estimated from measured weight and height.

Because many of these covariates are time dependent, we used a measure that represented the entire exposure period (1991-2006). For income, physical activity, urbancity index and BMI a mean of all the repeated measures was calculated. For smoking and alcohol intake we estimated the proportion of the participating waves in which they reported being currently smoking or consuming three or more alcoholic beverages per week. Education was defined as the highest level of education attained during each subject's follow-up period. Age in 2006 (equivalent to using birth year), gender and region were not time-dependent covariates.

Statistical analysis

Dietary pattern

In a previous study we compared in 2006 dietary patterns derived from Principal Component Analysis and RRR (HbA1c, HOMA-IR and glucose in 2009 as response variables). As expected the RRR pattern had a stronger association with HbA1c, insulin resistance and diabetes prevalence, and this pattern also seemed to be behaviorally meaningful as it shared many characteristics with the Principal Component patterns. Details of development of this dietary pattern are presented elsewhere (Batis et al, unpublished data, 2013). Briefly, 29 food groups were included in the analysis. Because of the large proportion of non-consumers, intake was categorized as binary: non-consumers vs. consumers for food groups with <80% of consumers and below vs. above the median otherwise (rice, fresh non-leafy vegetables and fresh leafy vegetables). With the residuals method we adjusted the food groups by geographical region, urbancity index, education and income (residuals were estimated from a regression with the food group as the dependent variable and the demographic variables as the independent ones). RRR was performed on the residuals with PROC PLS and RRR option (SAS 9.3, SAS Institute Inc., Cary, NC) with log-transformed HbA1c, HOMA-IR and fasting glucose as response variables. As customary when using RRR diet patterns, we retained only the first pattern, which explained the most of the variation in the response variables.

Dietary pattern score from 1991 to 2006

To assess how much each subject adhered to the dietary pattern, we computed a dietary pattern score for each subject at each wave: the higher the score, the more closely the participant's diet conforms to the dietary pattern. The loadings of all food groups on the pattern previously found in 2006 were used to calculate the pattern scores in all other waves (1991-2004). We used the same loadings in all waves, so that changes on a score that represents the same dietary pattern across time could be assessed. To simplify the computation of the score and make it more interpretable over time, we did a weighted sum (using the loadings as weights) of the original binary food groups (instead of the residuals) [i.e. Score= $\text{rice}_{(0,1)}$*-0.21 + wheat noodles $_{(0,1)}$*0.24 + wheat $\text{flour}_{(0,1)}$*0.13 + wheat buns and $\text{breads}_{(0,1)}$*0.36 ….].

Latent class trajectory analysis (LCTA)

Once every individual had a dietary pattern score in each wave, we used LCTA performed in Mplus 6.1 (Muthén & Muthén, Los Angeles, California). This method identifies subgroups of people with similar trajectories on their dietary pattern's scores over time. Contrary to conventional growth model approaches, in which the trajectory of all individuals is described using a single estimate of growth parameters and the heterogeneity is captured only by variations in the slope and intercepts (random effects), LCTA allow each class to have different parameters to represent their trajectory.[35,36] It should be noted that, the related but more complex, growth mixture model can also estimate random effects within each class. However because this model can have more convergence problems and we were not interested in the variability within each class we used LCTA, in which no variation across individuals is allowed within classes.[35]

For the model fitting, the analyst specifies the number of classes desired and the functional form of the growth trajectories. We found that the observed mean trajectories were fairly linear, even when using an unspecified shape (factor loadings that describe the functional form of the trajectories are freely estimated). In addition, the way subjects were classified into classes was not different when specifying a linear versus an unspecified shape; therefore we chose the more parsimonious linear shape. Suggested criteria to decide the number of classes, include interpretability, better model fit (lower BIC); models that do not have classes with a low proportion of participants; and higher entropy (close to 1) which is a measure of classification accuracy based on the posterior probabilities once individuals have been assigned to their most likely class. We compared 1 to 7 classes and decided to retain 5 classes; the fit of the model was better between 5 and 6 classes, but the 6-class model had a class with only 1.5% of the sample. In addition, models with 2 to 4 classes yielded parallel trajectories, which are less interesting to compare than trajectories that intersect each other, because this means classes share a similar diet at that point. Entropy was very low in all models, it ranged from 0.72 (2 classes) to 0.57 (5 classes) (Supplemental Table 5.1).

For estimating each model, we specified 100 initial stage random starts, 10 final stage optimizations and 10 initial stage iterations. We re-run the 5 class model with the second best seed value to confirm that the estimates were replicated and the solution was not local.[35]

Association between trajectories and outcomes

To assess the relationship between the dietary patterns trajectories and the outcomes, we run multiple linear (for HbA1c and HOMA-IR) and logistic (for diabetes) regressions with class trajectory membership as an indicator variable. Because HbA1c and HOMA-IR were natural log-transformed, the regression coefficients were multiplied by 100 and interpreted as the percent change in the outcome for being in a given class compared to the reference class. First we adjusted by gender, smoking, alcohol, education, geographical region, age, income, urbaniciy index and physical activity. In a second model we additionally adjusted by BMI. For all time-dependent covariates we used a measured that represented the entire follow-up period as previously explained. The clustering at the household level was accounted for in the estimation of the variance. Except for the RRR and LCTA, all other analyses were conducted in Stata 12.1 (StataCorp, College Station, TX).

Results

The dietary pattern identified in 2006 from RRR with HbA1c, HOMA-IR and fasting glucose as response variables was positively associated with the intake of wheat noodles, wheat buns/breads, deep-fried wheat and soy milk; and was negatively associated with rice, fresh legumes, poultry/game, eggs and fish/seafood (Supplemental Table 5.2). The association of this pattern with HbA1c, HOMA-IR and diabetes prevalence was positive; in 2006, comparing to the lowest dietary pattern score quartile, the percent change in HbA1c for quartiles 2 to 4 was 2.07 (0.97, 3.16), 2.58 (1.61, 3.55) and 3.92 (2.87, 4.97) respectively. For HOMA-IR it was 2.42 (-4.14, 8.97), 5.04 (-1.95, 12.03) and 9.05 (1.79, 16.32); and the Odds Ratio for Diabetes comparing to the lowest dietary pattern score quartile were 1.75 (1.04, 2.92), 2.11 (1.28, 3.48) and 2.02 (1.23, 3.32).

We also confirmed that in this sample, most of the pattern's key food groups had an independent association with HbA1c consistent with the direction of their factor loading. The proportion of waves from 1991-2006 in which the food group was consumed was positively associated with HbA1c in 2009 for wheat noodles, buns/breads ($p<0.05$) and deep-fried wheat ($p>0.05$); and negatively associated for rice, poultry/game ($p<0.05$), fresh legumes, eggs, fish/seafood and soy milk ($p>0.05$). Therefore, only soy milk, which is frequently consumed with deep-fried wheat, was inconsistent, it had a positive loading in the dietary pattern but its independent association with HbA1c tended to be negative (Supplemental Table 5.3).Therefore, the score for this dietary pattern could be interpreted as a weighted sum of healthy and unhealthy food groups, with unhealthy foods having a positive weight (their intake increases the score) and healthy foods having a negative weight (their intake decreases the score). Subjects with total negative scores consumed more of the healthy foods relative to their intake of unhealthy ones, subjects with positive scores, consumed more of the unhealthy foods relative to their intake of healthy foods, and subjects with scores near zero had an intake of unhealthy foods comparable to that of healthy ones.

Using LCTA we identified 5 classes of subjects with similar trajectories of their dietary pattern score from 1991 to 2006. The score over time of each class is shown in Figure 5.1, classes 1 and 2 had an increase in the score over time, whereas classes 3, 4, 5 were parallel to each other and had a relatively stable score over time. To better understand what these trends in dietary pattern score represented in terms of intake at the food group level, in Table 5.1 we present the proportion of consumers and mean number of food groups consumed over time. In all trajectory classes the diversity of diet increased, as the number of food groups consumed increased from 1991 to 2006. In class 1, which had a slight increase in the dietary pattern score from 1991 to 2006, the proportion of consumers increased importantly for all the relevant unhealthy foods (those with positive loadings, such as wheat noodles, wheat buns/breads, deep-fried wheat and soy milk), whereas the intake of healthy foods increased only for fresh legumes and eggs but remained relatively stable for rice, poultry/game, fish/seafood. Class 2 had the most dramatic increase in the score. The subjects in this class decreased the intake of most healthy foods and increased the intake of all unhealthy foods over time. For example, the proportion consuming one unhealthy food -- wheat buns and breads -- increased from ≤3 % to nearly 70% in both classes 1 and 2. However, the proportion consuming one healthy food – legumes – declined only in class 2.

The majority of subjects were in classes with relatively constant scores over time, with 47% of the sample in class 4 and 30% in classes 3 and 5. In all of these classes, the intake of rice decreased from 1991 to 2006 whereas for all other relevant food groups the proportion of consumers increased. Because the increases in both healthy and unhealthy foods were similar, the balance between the intake of healthy and unhealthy, and hence the dietary pattern scores for these groups remained stable over time. In addition class 3, for which the dietary pattern scores were consistently higher than classes 4 and 5, had a consistently higher proportion of intakes of unhealthy foods (wheat noodle and wheat buns/breads), and a lower proportion of intakes of healthier foods (e.g. rice, poultry/game, fish/seafood).

When comparing the class with the lowest (class 5) and highest score (class 1), those with lower scores tended to have higher education, income, urbanicity index, alcohol intake, proportion of males and lower physical activity and smoking (Table 5.2). Although what we present for income and urbanicity level represent the mean from 1991-2006, both variables had increased over this period in all classes. Nonetheless, in 1991 subjects in class 2 had a similar urbanicity index to class 4 however the increase over time for both variables was less pronounced in class 2. In terms of education, class 2 had the highest education level since 1991 (data not shown). There was a strong regional difference among the classes, 80% of class1 lived in the Central region, probably related to their lower consumption of rice and the higher consumption of buns/breads, whereas only 1.5% of class 5 lived in the North.

The adjusted mean HbA1c for each class is shown in Figure 5.1, and the unadjusted and adjusted percent difference between each pair of classes is shown in Table 5.3. As expected, the higher the overall 1991-2006 dietary pattern score, the higher the HbA1c in

2009. Interestingly, class 1 and class 2 had a similar score in 2006 but for class 2, which had a lower mean score over time, the HbA1c was 1.64 (0.11, 3.17) percent lower. Similarly in 1991, class 2 and class 4 had a similar score, but HbA1c tended to be slightly higher for class 2 [1.02 (-0.30, 2.35)]. The smallest HbA1c difference [0.62 (-0.72, 1.96)] was seen among the classes that had the most similar mean score over time (class 2 and 3), even if one had a positive slope and the other had a flat one. Results for HOMA-IR and the odds of being newly diagnosed were comparable to those for HbA1c, although estimates were less significant. Adjusting by BMI brought all the estimates closer to the null.

Discussion

To the best of our knowledge this is the first time LCTA is used to study the health effect of different trajectories of a dietary pattern. We found that among two classes that had trajectories with similar scores in 2006, the one that previously had lower scores, and hence a healthier diet, had a lower HbA1c in 2009. This suggests that even if the diet scores were similar recently, the different dietary trajectories that subjects followed to get to that point might also affect HbA1c levels. In addition, when comparing two classes that cumulatively had a similar mean dietary pattern score over time, there was no difference in HbA1c, even if one class had a stable score and the other had an increase over time. This suggests that when the dietary pattern score is similar cumulatively, there might not a difference between keeping a constant dietary pattern score throughout the follow-up period versus having changes (i.e. increasing score).

Only few studies have looked at the association between dietary patterns and health outcomes with repeated measures of dietary intake. The types of methods available for analyzing repeated measures vary considerably in their approach and aims, and therefore more research is needed to understand which one is the most meaningful way to represent the long-term exposure to a dietary pattern. Several studies used longitudinal mixed models to incorporate several repeated measures of a dietary pattern; the goal was to obtain an average dietary pattern-health outcome association adjusted by the inter-individual correlation of repeated measures.[31,115] In another study, instead of obtaining an average effect from the repeated measures of a dietary pattern, the independent effect of the dietary pattern measured at each point on the outcome was assessed; this approach was used to evaluate how long the effect of a dietary pattern could persist and affect the outcome.[116] As can be seen none of the previous analysis evaluated the inter-individuals changes or cumulative exposure to dietary patterns. Among other approaches, one study evaluated, with obesity as the outcome, the sequence of transition between different clusters of dietary patterns in three time points (i.e. staying in the same dietary pattern, change to a healthier or unhealthier one, etc.).[112] Others looked at the change in dietary patterns' scores between two time points and its association with weight, BMI change and obesity.[20,30] Another study estimated for each individual their intercept and slope over three repeated dietary patterns scores and used the dietary pattern's intercept and slope as predictors of intelligence quotient during childhood and adolescence.[113] The limitation of the last two approaches is that they can only evaluate one at a time the effect of either the dietary score at baseline or the change (slope) during follow-up, when the effect of one might be related to level of the other.

In our analysis we applied LCTA, for dietary exposures, only a similar method, growth mixture model, has been previously used to identify trajectories of sodium adherence over a 6-month period in 279 adults with heart failure.[117] The advantage of this approach is that a single variable (trajectory class membership) encompasses both the starting point (intercept) and the change (slope) in dietary pattern score during follow-up. Furthermore, by encompassing the change (slope) and the level at any given time (i.e. intercept, midpoint or last point) it implicitly summarizes also the mean score of the follow-up. Therefore the trajectory class gives a comprehensive representation of the long-term exposure to a dietary pattern score. It is straightforward to examine the trajectories relation to the predictors and outcomes, as there is no need to separately examine the association with the intercept and slope for example. In the same lines, LCTA was useful for describing the individual trends over time in this dietary pattern. When the interest is not only on the mean trajectory of the sample but on the individual's heterogeneity in the trajectories, it is easier to get a picture of the trajectories distribution with the data-driven LCTA than by looking simultaneously at the descriptive statistics of both the intercept and the slope. In our sample we found the majority kept a moderate constant score over time, and that the individual variation was mainly in the intercept.

A disadvantage that comes with a data-driven method like LCTA is that the model might not identify trajectories that enable meaningful comparisons. As previously explained the trajectory classes summarize the intercept, slope, mean score of the follow-up and last score. Each of these aspects might have a different and independent effect on a health outcome but it is not possible to tease them apart. Therefore, when interpreting the differences between the trajectory classes it is unknown how each characteristic of the trajectory is contributing to the effect observed. It is possible to exclude the effect of one of these characteristics at a time (i.e. comparing trajectory classes with similar end point or similar mean scores); however the comparison options available depend on the identified trajectories. It is of particular interest to understand if the slope has an effect by itself, because otherwise just evaluating the score at each time independently or the mean score would be equally informative. One might hypothesize that even with a same mean; there are different effects for having a constant score, versus increasing or decreasing it over time. By comparing class 2 and 3, our results suggest that the slope was not associated among subjects with similar mean scores. But it should be noted that even in the case that the slope was found to be relevant, it still would not be possible to understand if this was due to slope or to the score at the intercept or last point.

Other limitations we found when applying LCTA on dietary data was that the entropy of our model was very low (0.57). This means that the uncertainty in trajectory membership was high, which is likely when using dietary data, due to its high random variability. For example the entropy found in the study of sodium adherence was 0.75 for a 2-class model,[117] which is comparable to what we found in our 2-class model (entropy= 0.72). In contrast, studies on trajectories of BMI have found entropies of 0.80-0.85 for a 5-class models.[118,119] Other disadvantage is that sample size can be small for some classes.

In terms of how meaningful were the differences in HbA1c between the trajectory classes, we found that although they might seem small (i.e. HbA1c for class 1 and 2 was 5.65 and 5.55 respectively), for HbA1c this is typical of what other studies have found. For instance, in one study those who had fruit and green leafy vegetables seldom or never had a HbA1c of 5.43 versus 5.34 for those with a more frequent consumption.[120] Also, for alcohol abstainers the HbA1c was 5.8 compared to 5.6 for those consuming 0.1-1.9 drinks/day.[121] And the range in HbA1c for quintiles of total fat or polyunsaturated/saturated fat ratio was around 5.2 to 5.4.[122] Although our results were significant and meaningful for HbA1c, this was not the case for HOMA-IR and diabetes prevalence. It is possible that dietary intake is less associated and explains less of the variation in HOMA-IR compared to HbA1c, and this could be partially due to the fact that HOMA-IR includes glucose information which is a shorter term measure of glucose control. Diabetes prevalence was defined using HbA1c, and accordingly the results for diabetes were in the same direction as for HbA1c; however they were not significant in general. It is possible that more power was needed, particularly because only 5.6% of the sample had diabetes and the analysis required information for many stratums of the sample.

We found that adjusting by BMI brought our estimates only slightly closer to the null, which suggests that the trajectories of our dietary pattern score were directly associated with HbA1c, independently of the increased energy intake and increased BMI, which may be on the causal pathway. Proposed mechanisms by which diet relates to diabetes independent of increased energy include modulation of oxidative stress and inflammation, which in turn affect insulin sensitivity and β-cells function.[70,123] Our dietary pattern score represented high intake of foods such as wheat noodles, breads and buns and low intake of fish and legumes. It has been suggested that high glycemic index foods by promoting hyperglycemia are pro-inflammatory and increase oxidative stress,[123] whereas n-3 polyunsaturated fatty acids are anti-inflammatory and might improve the physical properties of cellular membranes and the binding affinity of insulin receptors;[71] legumes, in turn, are source of antioxidants like phenolic acids and fiber, which also has anti-inflammatory properties.[90,123]

Even when a dietary pattern, compared to single nutrients or food groups, can approximate more the complexity of dietary intake, it is still a very limited and non-exhaustive representation of diet. Our dietary pattern score was a weighted sum of the intake of healthy (negative weights) and unhealthy foods (positive weights). Despite the marked increased in diet diversity in the entire sample, many trajectories had a stable score over time, which means that the score did not capture diet diversity but rather reflects the overall balance between the intake of healthy and unhealthy foods related to the outcomes of interest. It is biologically plausible that this balance represented by the score has an effect on HbA1c regardless of diet diversity; however whether diet diversity has an additional effect should be explored by future research. Moreover, our analysis and dataset might not be able to capture some key dietary changes. For instance, the information we used for food intake was only about whether the food was consume or not, and we missed the amount consumed for most of the food groups. Also because in 3-day average many key food groups such western-style fast food, salty snacks or calorically-sweetened beverages, were scarcely consumed they were not included in the analysis. In addition, as in most countries, the Chinese food composition table and hence our dietary assessment focuses on the measurement of foods as purchased at the ingredient level, and not as consumed at the dish level, so changes in food and dishes preparation were not captured. Despite the limitations in the dietary assessment the methodology of detailed collection during a 3-day period has remained unchanged in all the surveys; this is a key strength that enables us to assess the longitudinal effects of dietary intake. We were able to use an innovative method such as LCTA to evaluate trajectories of dietary patterns, over a 15-year period follow-up, for the first time.

Another limitation is that not all subjects had complete follow-up, fortunately LCTA can use all of the information available without the need of imputing values or deleting cases with incomplete data. Even individuals with only one wave of data can be classified in one trajectory class, although one point in time does not seem informative enough. As a way to balance both internal and external validity we included individuals with at least 3 waves of data complete. Finally, compared to the thorough and complex longitudinal modeling of our main exposure with LCTA, the modeling of time-varying confounders was simplified. Deriving and adjusting by trajectories of each confounder might not only be challenging, but it might also raise non-convergence or non-positivity issues. Still we cannot disregard the potential of residual confounding of our approach.

In sum, the long-term exposure to a dietary pattern seems to be relevant HbA1c at a given point in time. More research using longitudinal methods is needed to confirm and add evidence or insights about the nature of these long-term relations. Findings like this can have important intervention and clinical implications. Subjects can be counseled to make intensive diet and lifestyle changes; given the case that their previous diet has already increased their risk of disease. Or better informed interventions to promote an early start and maintenance of healthy eating patterns throughout adulthood life could be implemented.

Tables and Figures

Table 5.1 Number of food groups and proportion of consumers for food groups associated with dietary pattern (Factor loadings > |0.20|) by wave and trajectory class.

	Class 1			Class 2			Class 3			Class 4			Class 5		
	1991	2000	2006	1991	2000	2006	1991	2000	2006	1991	2000	2006	1991	2000	2006
Number of food groups	5.9	7.7	8.7	7.2	8.1	9.4	6.8	7.9	8.8	7.4	8.3	9.1	8.3	9.7	10.5
Proportion of consumers															
Food groups with positive factor loadings (≥0.20)															
Wheat noodles	0.12	0.43	0.63	0.15	0.47	0.71	0.39	0.53	0.57	0.21	0.37	0.48	0.06	0.23	0.30
Wheat buns, breads	0.03	0.39	0.69	0.00	0.21	0.68	0.03	0.14	0.35	0.01	0.06	0.17	0.00	0.01	0.05
Deep-fried wheat	0.06	0.19	0.27	0.05	0.11	0.30	0.06	0.06	0.11	0.04	0.03	0.08	0.04	0.03	0.11
Soy milk	0.01	0.08	0.15	0.00	0.04	0.21	0.01	0.03	0.08	0.00	0.02	0.06	0.00	0.01	0.06
Food groups with negative factor loadings (≤ -0.20)															
Rice*	0.08	0.13	0.10	0.78	0.60	0.40	0.64	0.56	0.52	0.92	0.77	0.67	0.95	0.78	0.61
Fresh legumes	0.23	0.36	0.45	0.65	0.47	0.45	0.27	0.50	0.60	0.38	0.49	0.57	0.60	0.71	0.68
Poultry and game	0.02	0.05	0.03	0.12	0.13	0.04	0.01	0.07	0.07	0.13	0.22	0.25	0.35	0.49	0.50
Eggs and eggs products	0.19	0.46	0.56	0.45	0.51	0.56	0.17	0.41	0.54	0.36	0.53	0.60	0.64	0.66	0.78
Fish and seafood	0.08	0.11	0.10	0.36	0.26	0.15	0.12	0.17	0.22	0.44	0.50	0.53	0.66	0.79	0.86

*Proportion below and above the median

Table 5.2 Means or percent for demographic characteristics by dietary pattern trajectory class

	Class 1	Class 2	Class 3	Class 4	Class 5
N (%)	745 (18.2)	245 (6.0)	994 (24.3)	1,911 (46.7)	201 (4.9)
Mean dietary pattern score	0.36 ± 0.14	0.00 ± 0.12	0.03 ± 0.09	-0.34 ± 0.14	-0.73 ± 0.11
Age in 2006 (years), mean ± SD	48.9 ± 9.5	47.9 ± 8.9	48.8 ± 9.5	47.9 ± 9.4	50.2 ± 9.7
Region, %					
South	5.2	42.5	43.4	64.6	41.3
Central	80.0	26.9	26.7	18.0	57.2
North	14.8	30.6	30.0	17.4	1.5
Male sex %	45.8	46.9	46.1	47.9	49.3
BMI (kg/m^2)[a], mean ± SD	23.4 ± 2.8	23.1 ± 3.0	22.8 ± 2.8	22.5 ± 2.8	22.3 ± 2.4
Highest level of education attained[b], %					
None	21.1	13.9	17.7	12.2	11.9
Primary school	28.2	20.0	29.1	25.9	25.9
≥Lower middle school	50.7	66.1	53.2	61.9	62.2
Income[c], %					
Low	46.4	24.1	35.8	27.5	10.5
Medium	29.1	38.4	36.2	35.0	21.4
High	24.4	37.6	28.0	37.6	68.2
Urbancity[c], %					
Low	47.8	33.5	37.3	27.5	6.5
Medium	27.9	33.5	37.0	34.3	39.3
High	24.3	33.1	25.7	38.3	54.2
Not currently smoking in all participating waves, %					
Female	92.8	93.9	89.9	94.9	100.0
Male	16.1	22.6	17.7	18.8	18.2
Alcohol intake <3 times/week in all participating waves, %					
Female	93.6	90.8	92.0	93.2	92.2
Male	42.5	35.7	37.3	43.3	32.3
Physical Activity[c], %					
Low	30.1	39.2	27.9	34.5	43.8
Medium	39.3	33.9	33.5	33.5	31.3
High	30.6	26.9	38.6	32.0	24.9

[a]Mean of participating waves from 1991-2006; [b]Highest level attained in participating waves until 2006; [c]Tertiles of the mean of participating waves between 1991-2006

Table 5.3 Association between dietary pattern trajectory classes from 1991-2006 and HbA1c, HOMA-IR and newly diagnosed diabetes prevalence in 2009

	Class 1	Class 2	Class 3	Class 4	Class 5
			HbA1c, % change (95% CI)		
Unadjusted	0	-5.34 (-6.81, -3.86)	-4.51 (-5.62, -3.41)	-7.41 (-8.47, -6.36)	-6.85 (-8.58, -5.11)
Adjusted model 1 (changing reference class)					
Reference class:1	0	-1.64 (-3.17, -0.11)	-1.02 (-2.18, 0.14)	-2.66 (-3.86, -1.46)	-5.24 (-6.93, -3.55)
Reference class:2	1.64 (0.11, 3.17)	0	0.62 (-0.72, 1.96)	-1.02 (-2.35, 0.30)	-3.60 (-5.44, -1.76)
Reference class:3	1.02 (-0.14, 2.18)	-0.62 (-1.96, 0.72)	0	-1.64 (-2.54, -0.74)	-4.22 (-5.80, -2.65)
Reference class:4	2.66 (1.46, 3.86)	1.02 (-0.30, 2.35)	1.64 (0.74, 2.54)	0	-2.58 (-4.09, -1.07)
Reference class:5	5.24 (3.55, 6.93)	3.60 (1.76, 5.44)	4.22 (2.65, 5.80)	2.58 (1.07, 4.09)	0
Adjusted model 2	0	-1.64 (-3.13, -0.15)	-0.83 (-1.95, 0.30)	-2.37 (-3.53, -1.21)	-4.32 (-6.00, -2.65)
			HOMA-IR, %change (95% CI)		
Unadjusted	0	-11.77 (-22.45, -1.10)	-13.85 (-20.94, -6.76)	-16.47 (-22.72, -10.21)	-5.22 (-15.66, 5.23)
Adjusted model 1	0	-6.47 (-17.37, 4.42)	-6.07 (-13.73, 1.58)	-9.74 (-17.31, -2.18)	-9.96 (-20.72, 0.79)
Adjusted model 2	0	-6.46 (-16.55, 3.62)	-4.45 (-11.64, 2.74)	-7.29 (-14.50, -0.09)	-2.23 (-12.64, 8.17)
			Diabetes, Odds Ratio (95% CI)		
Unadjusted	0	0.51 (0.28, 0.95)	0.61 (0.42, 0.88)	0.46 (0.33, 0.65)	0.36 (0.16, 0.79)
Adjusted model 1	0	0.86 (0.44, 1.67)	0.99 (0.67, 1.47)	0.96 (0.65, 1.42)	0.46 (0.20, 1.06)
Adjusted model 2	0	0.80 (0.40, 1.61)	1.03 (0.69, 1.54)	1.01 (0.68, 1.50)	0.60 (0.26, 1.40)

Adjusted model 1: age in 2006, gender, geographical region, mean income, mean urbanicity index, mean physical activity, proportion of waves smoking, proportion of waves with alcohol intake ≥3 times/week; Adjusted model 2: model 1 plus mean BMI; HbA1c and HOMA-IR were logarithmically transformed.

Supplemental Table 5.1 Model fit and trajectory classes characteristics

BIC	Sample-Size Adjusted BIC	Class	Proportion in each class	Intercept	Slope	Posterior probability	Entropy
23606.67	23581.25	1	1	-0.179	0.007	1.0	NA
19938.5	19903.55	1	0.41	0.059	0.014	0.90	0.72
		2	0.59	-0.347	0.001	0.93	
19545.42	19500.93	1	0.46	-0.176	0.006	0.77	0.60
		2	0.24	0.13	0.018	0.86	
		3	0.30	-0.432	-0.003	0.82	
19483.25	19429.24	1	0.17	0.154	0.02	0.82	0.61
		2	0.46	-0.336	0.001	0.79	
		3	0.33	-0.078	0.009	0.71	
		4	0.04	-0.582	-0.006	0.75	
19444.58	19381.03	1	0.18	0.151	0.021	0.78	0.57
		2	0.06	-0.341	0.036	0.56	
		3	0.24	0.026	-0.002	0.60	
		4	0.47	-0.323	-0.001	0.76	
		5	0.05	-0.592	-0.004	0.74	
19449.75	19376.67	1	0.04	-0.356	0.041	0.57	0.59
		2	0.01	-0.037	-0.041	0.63	
		3	0.06	-0.585	-0.004	0.72	
		4	0.18	0.155	0.021	0.78	
		5	0.46	-0.343	0.003	0.74	
		6	0.26	0.016	0	0.62	
19457.84	19375.23	1	0.05	-0.597	-0.004	0.72	0.58
		2	0.01	-0.703	0.075	0.68	
		3	0.45	-0.355	0.003	0.72	
		4	0.25	-0.157	0.016	0.56	
		5	0.01	-0.056	-0.037	0.62	
		6	0.19	0.15	0.022	0.78	
		7	0.05	0.136	-0.013	0.56	

BIC, Bayesian Information Criteria; NA, Not applicable

Supplemental Table 5.2 Dietary pattern derived from Reduced Rank Regression in 2006

Food groups	Factor loadings*
Rice	**-0.22**
Wheat noodles	**0.30**
Wheat flour	0.08
Wheat buns, breads	**0.46**
Cakes, cookies and pastries	0.00
Deep-fried wheat	**0.22**
Corn and coarse grain	0.09
Starchy roots and tubers	0.11
Fresh legumes	**-0.24**
Dried legumes	0.05
Legume products	0.03
Nuts and seeds	-0.11
Starchy roots products and tubers products	0.14
Fresh vegetables, non-leafy	0.01
Fresh vegetables, leafy	0.08
Pickled, salted or canned vegetables	-0.01
Dried vegetables	0.19
Fruits	-0.08
High-fat red meat	0.02
Low-fat pork	0.18
High-fat pork	-0.08
Processed meats	-0.15
Organ meats	-0.01
Poultry and game	**-0.37**
Eggs	**-0.23**
Fish and seafood	**-0.29**
Soy milk	**0.24**
Animal-based milk	-0.14
Instant noodles and frozen dumplings	0.09

*Factor loadings > |0.20| are in bold numbers

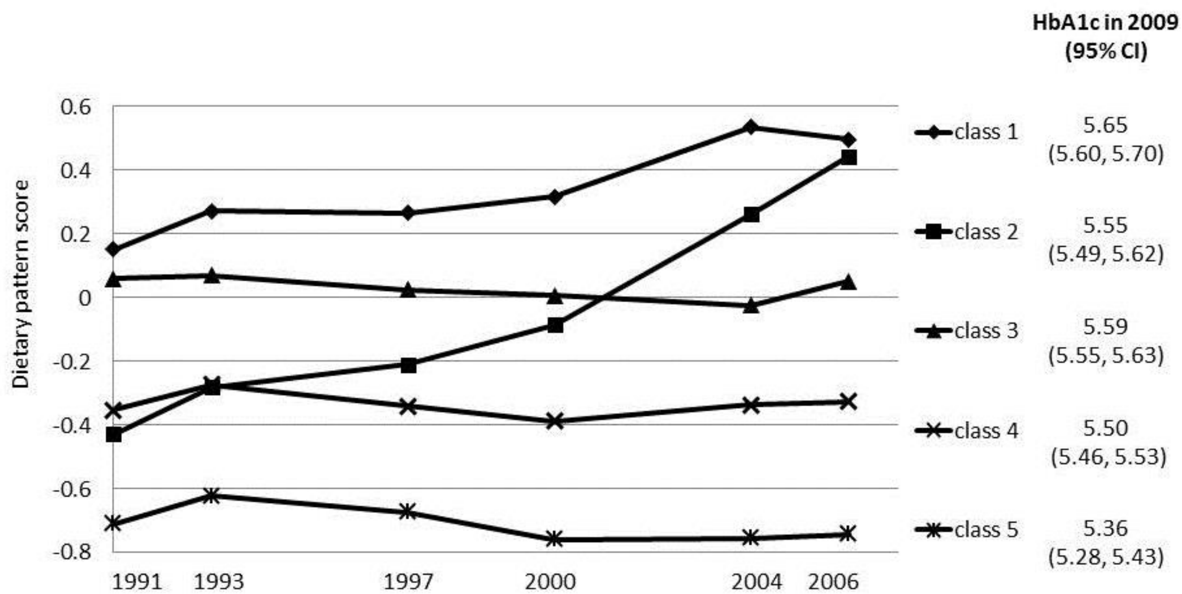
Supplemental Table 5.3 Independent association between relevant food groups and HbA1c.

	% change in HbA1c (95% CI)[a]
Rice[b]	-2.64 (-4.03, -1.25)
Wheat noodles	3.06 (1.75, 4.37)
Wheat buns, breads	2.50 (0.91, 4.08)
Deep-fried wheat	1.53 (-0.59, 3.66)
Fresh legumes	-0.64 (-1.98, 0.71)
Poultry and game	-1.95 (-3.86, -0.04)
Eggs and eggs products	-0.53 (-2.04, 0.98)
Fish and seafood	-1.20 (-2.54, 0.14)
Soy milk	-0.85 (-3.84, 2.13)

[a]For a 1 unit change in proportion of participating waves (1991-2006) the food group was consumed; i.e. the average difference in % change in HbA1c is 3.06 percentage points between those who always have consumed wheat noodles (proportion of consumption=1) vs. those who never had (proportion of consumption=0). Each coefficient was estimated in a different model adjusting by age in 2006, gender, geographical region, mean income, mean urbanicity index, mean physical activity, proportion of waves smoking, proportion of waves with alcohol intake ≥3 times/week.

[b]Proportion below and above the median, instead of proportion consumed.

Figure 5.1 Dietary pattern score trajectory classes from 1991 to 2006. The lines are the mean observed trajectory of each class; the mean HbA1c in 2009 adjusted by age, gender, geographical region, income, urbanicity index, physical activity, smoking, alcohol intake, is shown next to each class.



Chapter 6. Synthesis

Overview of findings

This research first focused on the dietary practice in China over the 1991-2009 period, how foods are eaten in combination, and how they changed or remained stable over time. This gave us insights about the longitudinal average dietary behaviors of the population. Then we examined whether these dietary patterns identified in the population were related to diabetes at one point in time only (2006). Finally we assessed if trajectories representing long-term dietary exposure were also related to diabetes and other glucose homeostasis biomarkers.

We used data from the China Health and Nutrition Survey (CHNS), a large and diverse survey with seven repeated detailed measurements of diet (1991, 1993, 1997, 2000, 2004, 2006 and 2009). The survey was designed to examine across space and time how economic and social shifts like modernization, are associated with a range of health behaviors. It has been conducted in China because of its unique rapidly changing economy and social system. All rounds of the CHNS have collected identical data from the community and household. Except for biomarkers, which were first measured in 2009. We focused on all adults 18 to 65 years old. Below we present a brief review of our findings for each aim.

1. Longitudinal Analysis of Dietary Patterns

Our purpose was to identify the changes or stability in the structure of dietary patterns and the tracking, trends and factors related to the adherence of these patterns in China from 1991 to 2009. We started by applying factor analysis in each wave to identify if the population's behavior, in term of which foods are eaten in combination, have changed or remained stable over time. We found that the structure of the dietary patterns have remained relatively stable, the patterns identified were a traditional southern pattern (rice, vegetables, meat, poultry and fish) and a modern high-wheat pattern (wheat products, nuts, fruits, eggs, milk and instant noodles/frozen dumpling).

Next, we calculated a dietary pattern score for each person in each wave to assess the level of adherence over time. We found that for the traditional southern dietary pattern there was less change in the relative position each subjects had, or in other words, that those that had high adherence at one point time maintained their high adherence over time; and those that had low adherence maintained it also. In addition, from the two dietary patterns identified only the modern high-wheat pattern showed an upward trend, meaning the popularity of this pattern increased. Geographical region, along with urbanicity level, income and education were the key predictors for both dietary patterns. Interestingly the differences in adherence between these subgroups of the population were becoming smaller over time, indicating that the reach of both dietary patterns might be spreading to the whole population.

Although many studies have identified dietary patterns in China, this is the first one to use repeated measures to evaluate the changes over time in these dietary patterns. Therefore this research filled an important gap in the literature.

2. Dietary patterns at one point in time and diabetes

The dietary patterns identified in the previous aim, are patterns that describe the eating behavior of the population, as the purpose of factor analysis is to find combinations of foods that explained the variation in the intake. However, the personal decisions for food selection might not be related to health. For example, the modern high-wheat dietary pattern did not only include unhealthful items such as cakes, cookies and pastries, deep-fried wheat, or instant noodles/frozen dumpling; but it also included healthy items such as nuts and fruits. Therefore, even if it is of interest to identify the health effects of these dietary patterns as they are, other methods such as Reduced Rank Regression (RRR), that aim to find the combination of foods that best explain the variation in the outcome are also useful when studying dietary pattern-disease associations. In this analysis we proposed that the use of both methods Principal Component Analysis (PCA) and RRR is complementary, and that the results of one are useful for the interpretation of the other.

Among 4,316 adults not previously diagnosed with diabetes, we found that the traditional southern dietary pattern in 2006 had a weak negative association with HbA1c, insulin resistance and diabetes prevalence in 2009, whereas the modern high-wheat pattern, had a weak positive association with these outcomes. As expected the dietary pattern identified with RRR had a strong positive association. Interestingly, the combination of foods that characterized the RRR pattern seemed to be behaviorally meaningful. This pattern probably combined the unhealthful aspects of the modern high-wheat (high intake of wheat buns & breads, deep-fried wheat, soy milk) with a low intake of the healthy items in the traditional southern (low intake of rice, poultry & game, fish & seafood). In addition fresh legumes was negatively associated with this pattern and wheat noodles were positively associated, indicating that even if these foods are not distinctive of the previous dietary patterns, they are key components of a pattern associated with diabetes in this population.

3. Long-term dietary patterns trajectories and diabetes

Most of the research that evaluates dietary pattern-disease associations use diet measured at only one point in time. They either assume that the individual intake or the ranking of their intake remains stable over time. However, individuals might have changes over time that do not line up with the average change in the population, and it is of interest to understand if there are any health effects associated with these inter-individual changes. Therefore, taking advantage of the repeated measures in the CHNS over 15 years of follow-up, we examined the long-term trajectories of the dietary pattern previously identified with RRR.

We used an innovative technique, not used before for dietary patterns to identify the trajectories. Latent Class Trajectory Analysis (LCTA) is a data-driven method that groups together subjects according to the similarities in their trajectories. The method was useful to describe the heterogeneity in the individual trajectories. In a model with 5 classes, we found that 3 classes had stable parallel scores over time, whereas two had positive slopes, one more pronounced than the other. The most interesting comparisons are between classes that intersect at some point. For example among two classes with similar scores in 2006, the one with lower scores previously had significantly lower HbA1c. And among two pair of classes with similar scores in 1991, the ones with increasing score over time had a slightly higher (non-significant) HbA1c.

After this analysis a key question still remained, are the trajectories important only because they summarize a long-term mean exposure, or is there an additional effect of changing (increasing or decreasing) versus maintaining a stable score over time. To complement the LCTA, we also estimated in each person their slope over time. We assessed the effect of the change in the slope while holding constant the mid-point of the trajectory, this way slopes were compared among subjects with similar cumulative mean over time, in this analysis we found that the slope had no effect. Therefore, these results suggest that the trajectory was important because it summarized the cumulative mean score and not because it represented patterns of change or stability.

Limitations and strengths

This research has several limitations; one of them is the potential for selection bias. The follow-up in the CHNS is complex because subjects might not be available in one wave and they come back later or new subjects are included in every wave as replenishment samples. Therefore it is not straightforward to identify which are the characteristics of the population our studies are trying to make inferences about, and therefore is hard to evaluate the severity of the potential selection bias. In every wave of the CHNS data for ~20,000 subjects is collected, but for instance in our first aim we included only 9,253 adults. Our inclusion criteria in this aims was to be aged 18 to 65 years old at any wave with at least 3 waves with complete dietary data from 1991 to 2007, only 20% had all 7 waves of diet complete. For our second aim only 4,316 were included in the analysis, the inclusion criteria was being 18-65 years old in 2006, not previously diagnosed with diabetes and have complete biomarker data in 2009. Furthermore, in the third aim, the analytic sample dropped even more to 4,049 subjects, they had the same characteristics of the sample of the second aim, except that instead of having diet data in 2006 they needed to have at least 3 waves of diet data complete. However, previous analysis conducted in the CHNS that have used methods to account for selection bias such as the Heckman two-equation approach or inverse probability weights, found that selectivity was not an issue.[124,125]

One important limitation in our data is that blood samples were collected for the first time in 2009; therefore biomarkers of glucose homeostasis were only known at the end of the follow-up. We were not able to distinguish between incident versus prevalent (but non-diagnosed) cases of diabetes. This was key for the longitudinal nature of our study, if subjects started having HbA1c $\geq$6.5% since 1991, then the dietary patterns from 1991 to 2006 was not the adequate time to evaluate the association between dietary intake and development of diabetes. On the other hand, even if subjects had HbA1c $\geq$6.5% since 1991, their diet intake could still played a role on their glucose control and hence on their HbA1c values, which is the main reason why in our analysis we did not focused only on a binary diabetes outcome, but also on the continuum of HbA1c values. Still, even when evaluating HbA1c values, it is possible that dietary intake has a different effect among subjects with prevalent diabetes over a longer period of time, compared to subjects that reached the HbA1c $\geq$6.5% cut point for the first time in 2009.

Another key limitation is related to the diet measurement. First of all, because it was measured during a short 3-day period it has more random-error and is less representative of usual intake, this could have weakened our estimates. Also most of the food groups had a low proportion of consumption; on one hand this increased the methodological complexity of our analysis, as most dietary pattern methods are developed for normally distributed variables. On the other hand we missed key information about diet intake. For instance by categorizing variables, particularly when categories were limited to consumers vs. non-consumer, we lost all the variation in amount consumed. For some food groups, which had <5 percent consumers, their intake was not captured at all in our analysis, this was the case of dairy products, candies/other high-sugar foods, western-style fast food, salty snacks or calorically-sweetened beverages. Unfortunately many of these items could be key indicators of diet modernization or key in the development of diabetes.

In addition, despite the short-term period the dietary assessment covered, another important limitation is that it is designed to capture intake of ingredients as purchased and not the intake of dishes as consumed. Food preparation is an important behavioral and nutritional aspect of diet intake, longitudinal changes in these might be occurring and we were not able to capture them. For example, we missed many elements that are indicative of modernization such as deep fried bread often consumed as a morning snack, if it was just coded as wheat and oil. Furthermore, cooking methods can alter the glycemic index of foods, and therefore be of particular relevance to the study of dietary patterns and diabetes.

Paradoxically, the main strength of this research lies also within the dietary assessment methodology. By combining both a 24-hr recall and household food inventory diet measurement is very detailed and precise. Also remarkable is the fact that in over almost 20 years of follow-up the methodology for dietary measurement has remained basically unchanged. Other common approaches, such as food frequency questionnaires, need to be designed and tailored specifically to the usual habits of the population in a given time. If used over a long follow-up they need to be updated to be able to capture the new trends in food intake. Therefore comparing intake over time, when assessed with different versions of a questionnaire is not straightforward, it is unclear whether the changes observed are real or due to the different questionnaire. The dietary assessment methodology in the CHNS has the great advantage of remaining constant while at the same time being capable of capturing the dietary intake changes that are evident in a 3-day period.

These seven repeated measures of dietary intake over almost 20 years of follow-up is the key strength of our analysis. In China there was no previous information about the longitudinal changes in dietary patterns. Moreover, in the general literature on dietary patterns, there has been only few exploration of how inter-individual changes relate to health outcomes. We used innovative longitudinal techniques to assess long-term exposure to dietary patterns. Characterizing dietary patterns trajectories with latent class trajectory analysis is an approach that has not been used in this particular area of research. Therefore with our analytical approach we were able to take full advantage of this unique data set.

Significance and public health impact

Our research has significant public health and research implications. On the research side, we contributed to the advancement of the field by thinking through its gaps and limitations and proposing new approaches for the study of dietary patterns. These new approaches were not only about the use of innovative methods, but also about a different use and interpretation of methods already established. On our second aim, we suggested that the use of both RRR and PCA is not only a matter of identifying which method is better, but about using the strengths of each to inform and help in the interpretation of the other. On our third aim we applied a method not previously used before for the study of dietary pattern. We outlined the advantages and disadvantages we found to inform other researchers of our experience, promoting a discussion about the utility or gains of using these methods. Our research not only brought substantive knowledge to the field, but also thoughts about the alternative methods and approaches that could be considered.

On the public health side, we described and documented the trends in dietary patterns of the Chinese population. We found that a dietary pattern characterized by the intake of wheat products has increased in its popularity; this pattern is associated with many energy dense foods but also with healthy items such as nuts and seeds or fruits. This information highlights the need of intervention among the increasing segment of the population following this pattern, so that the inclusion of healthy items becomes a more prominent part of their dietary pattern. In addition, the results of our second aim, indeed suggest that parts of this wheat-based dietary pattern is associated with diabetes when combined with low intake of healthy items such as legumes, fish and poultry.

Finally our research about the effects of long-term trajectories contributes to the understanding of the complex interplay between diet and disease. Individuals may have a similar dietary pattern at one point in time, but some of them could have recently adopted this pattern and some could have followed the same pattern for years. Our results suggest that the associated health risks for these two types of trajectories are different. If confirmed, this knowledge has clear direct public health applications. For instance subjects can be counseled to make intensive diet and lifestyle changes; given the case that their previous diet has already increased their risk of disease; or better informed interventions to promote an early start and maintenance of healthy eating patterns throughout adulthood life could be implemented.

Future directions

We proposed an innovative method to assess the long-term trajectories of dietary patterns and their effects on diabetes and markers of glucose homeostasis. More studies, in different populations and with different designs need to be conducted before the value of applying LCTA could be established. We found that the entropy or the accuracy of classification of subjects in trajectory classes was very low; it should be confirmed if this is the case for all dietary data, or if dietary measures that are more representative of usual intake would have a better performance with this method. Alternatively, it could be confirmed that despite the misclassification and low entropy, this technique is still useful. We found significant difference between trajectories for HbA1c, but not for diabetes prevalence. It is possible that identifying statistical differences among trajectories requires more power than differences between dietary patterns measured at one point in time only. Therefore, more studies with larger sample size are needed to identify if there is no association for diabetes prevalence or if it was a matter of lack of power.

As described previously, a key limitation in our research was that we were not able to distinguish between prevalent and incident diabetes. It should be a priority for the CHNS to keep collecting blood samples in all of the future waves, so that analyses like ours do not have this limitation. In addition, our results need to be confirmed in other surveys where available repeated measures of markers of glucose homeostasis are available.

There are still a lot of limitations and gaps in the measurement of dietary intake, particularly when the goal is to capture changes over time. Dietary measurement is the base for all subsequent analysis interested in dietary intake and resources need to be devoted for the improvement of this area. Methods need to be comparable over time, have a good approximation to usual intake and be able to capture changes in the food supply and preparation methods. A specific recommendation for the CHNS is to invest efforts in updating the food composition table, so that the introduction of new foods with new formulations could be better captured. Also, to include information about which ingredients reported in the 24-hr recall were consumed together in a dish with the corresponding dish name. We do not recommend however to replace the ingredient level by the dish level, because variations in recipes may introduce more error, and it is better to have the flexibility to aggregate or disaggregate the ingredients according to the specific interests of the researcher. Finally, the inclusion of short food frequency questionnaires, particularly for more episodically consumed foods, will be a useful complement to the dietary assessment. Statistical techniques could be used to incorporate information from the food frequency so that the measures from the 24-hour recall are more representative of usual intake.

Our analysis was very thorough about the changes of dietary patterns and their association with diabetes. However, the prevalence of diabetes is increasing dramatically in China and longitudinal analysis such as ours are needed for other modifiable risk factors. One key factor, worth of further analysis, is physical activity. In this population, physical activity, particularly occupational physical activity, had a very pronounced decreased over time and the changes in these were already associated with weight gain.[126-128] Hence, we recommend that future work address the longitudinal role of physical activity on diabetes, this will complement and help situate our findings in a broader context.

References

1. World Health Organization. *Diet, nutrition and the prevention of chronic diseases: report of a joint WHO/FAO expert consultation.* Geneva, Switzerland2002.

2. Popkin BM, Horton S, Kim S, Mahal A, Shuigao J. Trends in diet, nutritional status, and diet-related noncommunicable diseases in China and India: the economic costs of the nutrition transition. *Nutr Rev.* Dec 2001;59(12):379-390.

3. Popkin BM. The nutrition transition and obesity in the developing world. *J Nutr.* Mar 2001;131(3):871S-873S.

4. Yang ZJ, Liu J, Ge JP, et al. Prevalence of cardiovascular disease risk factor in the Chinese population: the 2007-2008 China National Diabetes and Metabolic Disorders Study. *Eur Heart J.* Jun 30.

5. Colin Bell A, Adair LS, Popkin BM. Ethnic differences in the association between body mass index and hypertension. *Am J Epidemiol.* Feb 15 2002;155(4):346-353.

6. Chan JC, Malik V, Jia W, et al. Diabetes in Asia: epidemiology, risk factors, and pathophysiology. *JAMA.* May 27 2009;301(20):2129-2140.

7. Deurenberg P, Deurenberg-Yap M, Guricci S. Asians are different from Caucasians and from each other in their body mass index/body fat per cent relationship. *Obes Rev.* Aug 2002;3(3):141-146.

8. Reynolds K, Gu D, Whelton PK, et al. Prevalence and risk factors of overweight and obesity in China. *Obesity (Silver Spring).* Jan 2007;15(1):10-18.

9. Du S, Lu B, Zhai F, Popkin BM. A new stage of the nutrition transition in China. *Public Health Nutr.* Feb 2002;5(1A):169-174.

10. Popkin BM, Lu B, Zhai F. Understanding the nutrition transition: measuring rapid dietary changes in transitional countries. *Public Health Nutr.* Dec 2002;5(6A):947-953.

11. Popkin BM. Will China's nutrition transition overwhelm its health care system and slow economic growth? *Health Aff (Millwood).* Jul-Aug 2008;27(4):1064-1076.

12. Moeller SM, Reedy J, Millen AE, et al. Dietary patterns: challenges and opportunities in dietary patterns research an Experimental Biology workshop, April 1, 2006. *J Am Diet Assoc.* Jul 2007;107(7):1233-1239.

13. Hoffmann K, Schulze M, Boeing H, Altenburg H. Dietary patterns: report of an international workshop. *Public Health Nutrition.* 2002;5(1):89-90.

14. Hu FB. Dietary pattern analysis: a new direction in nutritional epidemiology. *Curr Opin Lipidol.* Feb 2002;13(1):3-9.

15. Guenther PM, Krebs-Smith SM, Reedy J, et al. Healthy eating index-2005. *Center for Nutrition Policy and Promotion factsheet.* 2006(1).

16. Newby PK, Tucker KL. Empirically derived eating patterns using factor or cluster analysis: a review. *Nutr Rev.* May 2004;62(5):177-203.

17. Kant AK. Dietary patterns and health outcomes. *J Am Diet Assoc.* Apr 2004;104(4):615-635.

18. Togo P, Osler M, Sørensen T, Heitmann B. Food intake patterns and body mass index in observational studies. *International journal of obesity and related metabolic disorders: journal of the International Association for the Study of Obesity.* 2001;25(12):1741.

19. Newby P, Muller D, Hallfrisch J, Andres R, Tucker KL. Food patterns measured by factor analysis and anthropometric changes in adults. *The American journal of clinical nutrition.* 2004;80(2):504.

20. Newby PK, Weismayer C, Akesson A, Tucker KL, Wolk A. Longitudinal changes in food patterns predict changes in weight and body mass index and the effects are greatest in obese women. *J Nutr.* Oct 2006;136(10):2580-2587.

21. Esmaillzadeh A, Kimiagar M, Mehrabi Y, Azadbakht L, Hu FB, Willett WC. Dietary patterns, insulin resistance, and prevalence of the metabolic syndrome in women. *The American journal of clinical nutrition.* 2007;85(3):910.

22. Fung TT, Rimm EB, Spiegelman D, et al. Association between dietary patterns and plasma biomarkers of obesity and cardiovascular disease risk. *The American journal of clinical nutrition.* 2001;73(1):61-67.

23. Odegaard AO, Koh WP, Butler LM, et al. Dietary Patterns and Incident Type 2 Diabetes in Chinese Men and Women. *Diabetes Care.* 2011;34(4):880-885.

24. Fung TT, Schulze M, Manson JAE, Willett WC, Hu FB. Dietary patterns, meat intake, and the risk of type 2 diabetes in women. *Archives of internal medicine.* 2004;164(20):2235.

25. Montonen J, Knekt P, Härkänen T, et al. Dietary patterns and the incidence of type 2 diabetes. *American journal of epidemiology.* 2005;161(3):219.

26. Hodge AM, English DR, O'Dea K, Giles GG. Dietary patterns and diabetes incidence in the Melbourne Collaborative Cohort Study. *American journal of epidemiology.* 2007;165(6):603.

27. Crozier SR, Robinson SM, Godfrey KM, Cooper C, Inskip HM. Women's dietary patterns change little from before to during pregnancy. *J Nutr.* Oct 2009;139(10):1956-1963.

28. Mishra GD, McNaughton SA, Bramwell GD, Wadsworth ME. Longitudinal changes in dietary patterns during adult life. *Br J Nutr.* Oct 2006;96(4):735-744.

29. Newby PK, Weismayer C, Akesson A, Tucker KL, Wolk A. Long-term stability of food patterns identified by use of factor analysis among Swedish women. *J Nutr.* Mar 2006;136(3):626-633.

30. Togo P, Osler M, Sorensen TI, Heitmann BL. A longitudinal study of food intake patterns and obesity in adult Danish men and women. *International journal of obesity and related metabolic disorders.* Apr 2004;28(4):583-593.

31. McNaughton SA, Mishra GD, Stephen AM, Wadsworth ME. Dietary patterns throughout adult life are associated with body mass index, waist circumference, blood pressure, and red cell folate. *J Nutr.* Jan 2007;137(1):99-105.

32. Cutler GJ, Flood A, Hannan P, Neumark-Sztainer D. Major patterns of dietary intake in adolescents and their stability over time. *J Nutr.* Feb 2009;139(2):323-328.

33. Li J, Wang Y. Tracking of dietary intake patterns is associated with baseline characteristics of urban low-income African-American adolescents. *J Nutr.* Jan 2008;138(1):94-100.

34. Shi Z, Hu X, Yuan B, et al. Vegetable-rich food pattern is related to obesity in China. *Int J Obes (Lond).* Jun 2008;32(6):975-984.

35. Jung T, Wickrama KAS. An Introduction to Latent Class Growth Analysis and Growth Mixture Modeling. *Social and Personality Psychology Compass.* 2008;2(1):302-317.

36. Nagin DS, Odgers CL. Group-based trajectory modeling in clinical research. *Annu Rev Clin Psychol.* Apr 27;6:109-138.

37. Ostbye T, Malhotra R, Landerman LR. Body mass trajectories through adulthood: results from the National Longitudinal Survey of Youth 1979 Cohort (1981-2006). *Int J Epidemiol.* Feb;40(1):240-250.

38. Cloninger CR. A systematic method for clinical description and classification of personality variants. A proposal. *Arch Gen Psychiatry.* Jun 1987;44(6):573-588.

39. Kasen S, Cohen P, Skodol AE, Johnson JG, Smailes E, Brook JS. Childhood depression and adult personality disorder: alternative pathways of continuity. *Arch Gen Psychiatry.* Mar 2001;58(3):231-236.

40. Moffitt TE. Adolescence-limited and life-course-persistent antisocial behavior: a developmental taxonomy. *Psychol Rev.* Oct 1993;100(4):674-701.

41. Wang Y, Mi J, Shan X, Wang QJ, Ge K. Is China facing an obesity epidemic and the consequences? The trends in obesity and chronic disease in China. *International journal of obesity.* 2006;31(1):177-188.

42. Yang ZJ, Liu J, Ge JP, Chen L, Zhao ZG, Yang WY. Prevalence of cardiovascular disease risk factor in the Chinese population: the 2007–2008 China National Diabetes and Metabolic Disorders Study. *European heart journal.* 2012;33(2):213-220.

43. Wang Z, Zhai F, Du S, Popkin B. Dynamic shifts in Chinese eating behaviors. *Asia Pac J Clin Nutr.* 2008;17(1):123-130.

44. Wang Z, Zhai F, Zhang B, Popkin BM. Trends in Chinese snacking behaviors and patterns and the social-demographic role between 1991and 2009. *Asia Pacific Journal of Clinical Nutrition.* 2012;21(2):253.

45. Zhai F, Wang H, Du S, et al. Prospective study on nutrition transition in China. *Nutrition reviews.* 2009;67:S56-S61.

46. Wang D, He Y, Li Y, et al. Dietary patterns and hypertension among Chinese adults: a nationally representative cross-sectional study. *BMC Public Health.* 2011;11(1):925.

47. Li Y, He Y, Lai J, et al. Dietary patterns are associated with stroke in Chinese adults. *The Journal of nutrition.* 2011;141(10):1834-1839.

48. Zhang X, Dagevos H, He Y, Van der Lans I, Zhai F. Consumption and corpulence in China: a consumer segmentation study based on the food perspective. *Food Policy.* 2008;33(1):37-47.

49. Popkin BM, Du S, Zhai F, Zhang B. Cohort Profile: The China Health and Nutrition Survey--monitoring and understanding socio-economic and health change in China, 1989-2011. *Int J Epidemiol.* Dec 2010;39(6):1435-1440.

50. Bollen KA. *Structural Equations with Latent Variables*. New York: Wiley; 1989.

51. Maydeu-Olivares A, GarcÃ-a-Forero C, Gallardo-Pujol D, Renom J. Testing categorized bivariate normality with two-stage polychoric correlation estimates. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences.* 2009;5(4):131.

52. Kim JO, Mueller CW. *Factor analysis: Statistical methods and practical issues.* Vol 14: Sage Publications, Incorporated; 1978.

53. Northstone K, Emmett PM. A comparison of methods to assess changes in dietary patterns from pregnancy to 4 years post-partum obtained using principal components analysis. *British Journal of Nutrition.* 2008;99(05):1099-1106.

54. Elstgeest LEM, Mishra GD, Dobson AJ. Transitions in Living Arrangements Are Associated with Changes in Dietary Patterns in Young Women. *The Journal of nutrition.* 2012;142(8):1561-1567.

55. Muthén BO. *Mplus Technical Appendices*. Los Angeles: CA: Muthén & Muthén; 1998-2004.

56. HernÃ¡n MA, Robins JM. Estimating causal effects from epidemiological data. *Journal of Epidemiology and Community Health.* 2006;60(7):578-586.

57. Zhang Y, Tan H, Dai X, Huang H, He G. Dietary patterns are associated with weight gain in newlyweds: findings from a cross-sectional study in Shanghai, China. *Public Health Nutrition.*1(1):1-9.

58. Villegas R, Yang G, Gao YT, et al. Dietary patterns are associated with lower incidence of type 2 diabetes in middle-aged women: the Shanghai Womenâ€™s Health Study. *International journal of epidemiology*.39(3):889-899.

59. Cui X, Dai Q, Tseng M, Shu XO, Gao YT, Zheng W. Dietary patterns and breast cancer risk in the shanghai breast cancer study. *Cancer Epidemiology Biomarkers & Prevention.* 2007;16(7):1443-1448.

60. Cai H, Zheng W, Xiang YB, et al. Dietary patterns and their correlates among middle-aged and elderly Chinese men: a report from the Shanghai Men's Health Study. *British Journal of Nutrition.* 2007;98(5):1006-1013.

61. Zuo H, Shi Z, Yuan B, et al. Dietary patterns are associated with insulin resistance in Chinese adults without known diabetes. *The British journal of nutrition.*1.

62. Nanri A, Shimazu T, Ishihara J, et al. Reproducibility and Validity of Dietary Patterns Assessed by a Food Frequency Questionnaire Used in the 5-Year Follow-Up Survey of the Japan Public Health Center-Based Prospective Study. *Journal of Epidemiology*.22(3):205-215.

63. Hu FB, Rimm E, Smith-Warner SA, et al. Reproducibility and validity of dietary patterns assessed with a food-frequency questionnaire. *Am J Clin Nutr.* Feb 1999;69(2):243-249.

64. Khani BR, Ye W, Terry P, Wolk A. Reproducibility and validity of major dietary patterns among Swedish women assessed with a food-frequency questionnaire. *J Nutr.* Jun 2004;134(6):1541-1545.

65. Ng SW, Popkin BM. Monitoring foods and nutrients sold and consumed in the United States: Dynamics and Challenges. *Journal of the Academy of Nutrition and Dietetics.*112(1):41.

66. Hu D, Reardon T, Rozelle S, Timmer P, Wang H. The emergence of supermarkets with Chinese characteristics: challenges and opportunities for China's agricultural development. *Development Policy Review.* 2004;22(5):557-586.

67. Who EC. Appropriate body-mass index for Asian populations and its implications for policy and intervention strategies. *Lancet.* 2004;363(9403):157.

68. Jones-Smith JC, Popkin BM. Understanding community context and adult health changes in China: Development of an urbanicity scale. *Social Science & Medicine.*71(8):1436-1446.

69. Pan X-R, Yang W-Y, Li G-W, Liu J. Prevalence of diabetes and its risk factors in China, 1994. *Diabetes Care.* 1997;20(11):1664-1669.

70. Salas-Salvadó J, Martinez-Gonzalez MA, Bulló M, Ros E. The role of diet in the prevention of type 2 diabetes. *Nutrition, Metabolism and Cardiovascular Diseases.*21:B32-B48.

71. Thomas T, Pfeiffer AFH. Foods for the prevention of diabetes: how do they work? *Diabetes/metabolism research and reviews.*28(1):25-49.

72. Hu FB. Globalization of Diabetes The role of diet, lifestyle, and genes. *Diabetes Care.*34(6):1249-1257.

73. Michels KB, Schulze MB. Can dietary patterns help us detect diet-disease associations? *Nutrition research reviews.* 2005;18(2):241-248.

74. Hoffmann K, Schulze MB, Schienkiewitz A, Nothlings U, Boeing H. Application of a new statistical method to derive dietary patterns in nutritional epidemiology. *American journal of epidemiology.* 2004;159(10):935-944.

75. He Y, Ma G, Zhai F, et al. Dietary patterns and glucose tolerance abnormalities in Chinese adults. *Diabetes Care.* 2009;32(11):1972-1976.

76. Matthews DR, Hosker JP, Rudenski AS, Naylor BA, Treacher DF, Turner RC. Homeostasis model assessment: insulin resistance and Î²-cell function from fasting plasma glucose and insulin concentrations in man. *Diabetologia.* 1985;28(7):412-419.

77. International Expert C. International Expert Committee report on the role of the A1C assay in the diagnosis of diabetes. *Diabetes Care.* Jul 2009;32(7):1327-1334.

78. Hare MJ, Shaw JE, Zimmet PZ. Current controversies in the use of haemoglobin A1c. *Journal of internal medicine.* 2012;271(3):227-236.

79. Xin Z, Yuan M-X, Li H-X, et al. Evaluation for Fasting and 2-hour Glucose and HbA1c for Diagnosing Diabetes Based on Prevalence of Retinopathy in a Chinese Population. *PloS one.*7(7):e40610.

80. Yang C, Liu Y, Li X, Liang H, Jiang X. Utility of hemoglobin A1c for the identification of individuals with diabetes and prediabetes in a Chinese high risk population. *Scandinavian Journal of Clinical & Laboratory Investigation.* 2012;72(5):403-409.

81. Yu Y, Ouyang X-J, Lou Q-L, et al. Validity of Glycated Hemoglobin in Screening and Diagnosing Type 2 Diabetes Mellitus in Chinese Subjects. *The Korean journal of internal medicine.* 2012;27(1):41-46.

82. Ng SW, Norton EC, Popkin BM. Why have physical activity levels declined among Chinese adults? Findings from the 1991-2006 China Health and Nutrition Surveys. *Soc Sci Med.* Apr 2009;68(7):1305-1314.

83. Tucker KL. Dietary patterns, approaches, and multicultural perspective. *Appl Physiol Nutr Metab.* Apr;35(2):211-218.

84. McNaughton SA, Mishra GD, Brunner EJ. Dietary patterns, insulin resistance, and incidence of type 2 diabetes in the Whitehall II Study. *Diabetes Care.* 2008;31(7):1343-1348.

85. Imamura F, Lichtenstein AH, Dallal GE, Meigs JB, Jacques PF. Generalizability of dietary patterns associated with incidence of type 2 diabetes mellitus. *The American journal of clinical nutrition.* 2009;90(4):1075-1083.

86. Villegas R, Liu S, Gao Y-T, et al. Prospective study of dietary carbohydrates, glycemic index, glycemic load, and incidence of type 2 diabetes mellitus in middle-aged Chinese women. *Archives of internal medicine.* 2007;167(21):2310.

87. Villegas R, Gao YT, Yang G, et al. Legume and soy food intake and the incidence of type 2 diabetes in the Shanghai Women's Health Study. *Am J Clin Nutr.* Jan 2008;87(1):162-167.

88. Willett W, Manson J, Liu S. Glycemic index, glycemic load, and risk of type 2 diabetes. *The American journal of clinical nutrition.* 2002;76(1):274S-280S.

89. Sluijs I, van der Schouw YT, Spijkerman AM, Hu FB, Grobbee DE, Beulens JW. Carbohydrate quantity and quality and risk of type 2 diabetes in the European Prospective Investigation into Cancer and Nutritionâ€“Netherlands (EPIC-NL) study. *The American journal of clinical nutrition.*92(4):905-911.

90. Weickert MO, Pfeiffer AFH. Metabolic effects of dietary fiber consumption and prevention of diabetes. *The Journal of nutrition.* 2008;138(3):439-442.

91. Xun P, He K. Fish Consumption and Incidence of Diabetes Meta-analysis of data from 438,000 individuals in 12 independent prospective cohorts with an average 11-year follow-up. *Diabetes Care.*35(4):930-938.

92. Zheng J-S, Huang T, Yang J, Fu Y-Q, Li D. Marine N-3 Polyunsaturated Fatty Acids Are Inversely Associated with Risk of Type 2 Diabetes in Asians: A Systematic Review and Meta-Analysis. *PloS one.*7(9):e44525.

93. Villegas R, Shu XO, Gao Y-T, et al. The association of meat intake and the risk of type 2 diabetes may be modified by body weight. *International journal of medical sciences.* 2006;3(4):152.

94. Shi Z, Yuan B, Zhang C, Zhou M, Holmboe-Ottesen G. Egg consumption and the risk of diabetes in adults, Jiangsu, China. *Nutrition.*27(2):194-198.

95. Hu EA, Pan A, Malik V, Sun Q. White rice consumption and risk of type 2 diabetes: meta-analysis and systematic review. *BMJ: British Medical Journal.*344.

96. Zhang G, Pan A, Zong G, et al. Substituting white rice with brown rice for 16 weeks does not substantially affect metabolic risk factors in middle-aged Chinese men and women with diabetes or a high risk for diabetes. *The Journal of nutrition.*141(9):1685-1690.

97. Schulze MB, Hoffmann K, Manson JE, et al. Dietary pattern, inflammation, and incidence of type 2 diabetes in women. *The American journal of clinical nutrition.* 2005;82(3):675-684.

98. Heidemann C, Hoffmann K, Spranger J, et al. A dietary pattern protective against type 2 diabetes in the European Prospective Investigation into Cancer and Nutrition (EPIC) Potsdam Study cohort. *Diabetologia.* 2005;48(6):1126-1134.

99. DiBello JR, Kraft P, McGarvey ST, Goldberg R, Campos H, Baylin A. Comparison of 3 methods for identifying dietary patterns associated with risk of disease. *American journal of epidemiology.* 2008;168(12):1433-1443.

100. Nettleton JA, Steffen LM, Schulze MB, et al. Associations between markers of subclinical atherosclerosis and dietary patterns derived by principal components analysis and reduced rank regression in the Multi-Ethnic Study of Atherosclerosis (MESA). *The American journal of clinical nutrition.* 2007;85(6):1615-1625.

101. Manios Y, Kourlaba G, Grammatikaki E, Androutsos O, Ioannou E, Roma-Giannikou E. Comparison of two methods for identifying dietary patterns associated with obesity in preschool children: the GENESIS study. *European journal of clinical nutrition.*64(12):1407-1414.

102. Hoffmann K, Boeing H, Boffetta P, et al. Comparison of two statistical approaches to predict all-cause mortality by dietary patterns in German elderly subjects. *British Journal of Nutrition.* 2005;93(5):709-716.

103. Vujkovic M, Steegers EA, Looman CW, Ocka MC, van der Spek PJ, Steegersâ€ • Theunissen RgP. The maternal Mediterranean dietary pattern is associated with a reduced risk of spina bifida in the offspring. *BJOG: An International Journal of Obstetrics & Gynaecology.* 2009;116(3):408-415.

104. Zhai F, Guo X, Popkin BM, et al. Evaluation of the 24-hour individual recall method in China. *FOOD AND NUTRITION BULLETIN-UNITED NATIONS UNIVERSITY-.* 1996;17:154-161.

105. Kleiman S, Ng SW, Popkin B. Drinking to our health: can beverage companies cut calories while maintaining profits? *obesity reviews.*13(3):258-274.

106. Odegaard AO, Koh W-P, Arakawa K, Mimi CY, Pereira MA. Soft Drink and Juice Consumption and Risk of Physician-diagnosed Incident Type 2 Diabetes The Singapore Chinese Health Study. *American journal of epidemiology.*171(6):701-708.

107. Yang W, Lu J, Weng J, et al. Prevalence of diabetes among men and women in China. *New England Journal of Medicine.* 2010;362(12):1090-1101.

108. Yan S, Li J, Li S, et al. The expanding burden of cardiometabolic risk in China: the China Health and Nutrition Survey. *Obesity Reviews.* 2012;13(9):810-821.

109. Fung TT, Schulze M, Manson JE, Willett WC, Hu FB. Dietary patterns, meat intake, and the risk of type 2 diabetes in women. *Arch Intern Med.* Nov 8 2004;164(20):2235-2240.

110. Fung TT, Willett WC, Stampfer MJ, Manson JE, Hu FB. Dietary patterns and the risk of coronary heart disease in women. *Arch Intern Med.* Aug 13-27 2001;161(15):1857-1862.

111. Newby PK, Muller D, J H, Qiao N, Andres R, Tucker KL. Dietary patterns and changes in body mass index and waist circumference in adults. *The American journal of clinical nutrition.* 2003;77:1417-1425.

112. Pachucki MA. Food pattern analysis over time: unhealthful eating trajectories predict obesity. *International journal of obesity.* 2011;36(5):686-694.

113. Smithers LG, Golley RK, Mittinty MN, et al. Do Dietary Trajectories between Infancy and Toddlerhood Influence IQ in Childhood and Adolescence? Results from a Prospective Birth Cohort Study. *PloS one.* 2013;8(3):e58904.

114. Slining MM, Herring AH, Popkin BM, Mayer-Davis EJ, Adair LS. Infant BMI trajectories are associated with young adult body composition. *Journal of Developmental Origins of Health and Disease.* 2013;4(01):56-68.

115. Mikkilä V, Räsänen L, Raitakari OT, et al. Major dietary patterns and cardiovascular risk factors from childhood to adulthood. The Cardiovascular Risk in Young Finns Study. *British Journal of Nutrition.* 2007;98(01):218-225.

116. Northstone K, Joinson C, Emmett P, Ness A, Paus T. Are dietary patterns in childhood associated with IQ at 8 years of age? A population-based cohort study. *Journal of epidemiology and community health.* 2012;66(7):624-628.

117. Creber RM, Lee CS, Lennie TA, Topaz M, Riegel B. Using Growth Mixture Modeling to Identify Classes of Sodium Adherence in Adults With Heart Failure. *The Journal of cardiovascular nursing.* 2013.

118. Garden FL, Marks GB, Simpson JM, Webb KL. Body Mass Index (BMI) Trajectories from Birth to 11.5 Years: Relation to Early Life Food Intake. *Nutrients.* 2012;4(10):1382-1398.

119. Slining M, Herring A, Popkin B, Mayer-Davis E, Adair L. Infant BMI trajectories are associated with young adult body composition. *Journal of Developmental Origins of Health and Disease.* 2013;4(01):56-68.

120. Sargeant L, Khaw K, Bingham S, et al. Fruit and vegetable intake and population glycosylated haemoglobin levels: the EPIC-Norfolk Study. *European journal of clinical nutrition.* 2001;55(5):342.

121. Meyer KA, Conigrave KM, Chu N-F, et al. Alcohol consumption patterns and HbA1c, C-peptide and insulin concentrations in men. *Journal of the American College of Nutrition.* 2003;22(3):185-194.

122. Harding A-H, Sargeant LA, Welch A, et al. Fat Consumption and HbA1c Levels The EPIC-Norfolk Study. *Diabetes care.* 2001;24(11):1911-1916.

123. Avignon A, Hokayem M, Bisbal C, Lambert K. Dietary antioxidants: Do they have a role to play in the ongoing fight against abnormal glucose metabolism? *Nutrition.*

124. Jaacks LM, Gordon-Larsen, Penny, Mayer-Davis, Elizabeth J, Adair, Linda S., Popkin, Barry. Age and period effects on adult body mass index and overweight from 1991 to 2009 in China: The China Health and Nutrition Survey. *International Journal of Epidemiology.* in press.

125. Batis C, Gordon-Larsen P, Cole SR, Du S, Zhang B, Popkin B. Sodium Intake from Various Time Frames and Incident Hypertension Among Chinese Adults. *Epidemiology.* 2013;24(3).

126. Ng SW, Norton EC, Popkin BM. Why have physical activity levels declined among Chinese adults? Findings from the 1991-2006 China Health and Nutrition Surveys. *Social science & medicine.* Apr 2009;68(7):1305-1314.

127. Ng SW, Popkin BM. Time use and physical activity: a shift away from movement across the globe. *Obesity Reviews.* 2012:no-no.

128. Monda KL, Adair LS, Zhai F, Popkin BM. Longitudinal relationships between occupational and domestic physical activity patterns and body weight in China. *European journal of clinical nutrition.* Jul 18 2008;62(1318–1325).